

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2002年7月25日 (25.07.2002)

PCT

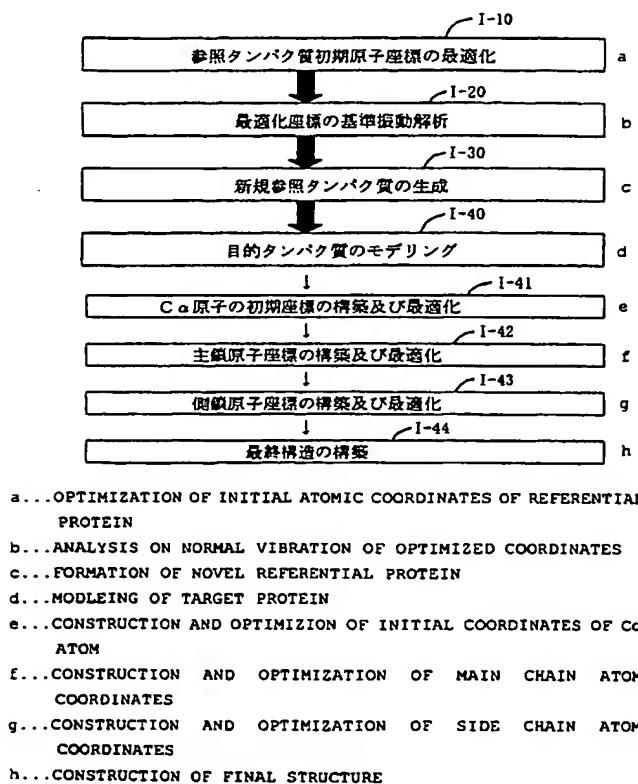
(10) 国際公開番号
WO 02/057954 A1

(51) 国際特許分類: G06F 17/30, 17/50, 19/00
 (71) 出願人および
 (21) 国際出願番号: PCT/JP02/00286
 (72) 発明者: 梅山秀明 (UMEYAMA, Hideaki) [JP/JP]; 〒279-0011 千葉県浦安市美浜1-7-1002 Chiba (JP).
 (22) 国際出願日: 2002年1月17日 (17.01.2002)
 (73) 発明者; および
 (25) 国際出願の言語: 日本語
 (75) 発明者/出願人(米国についてのみ): 小松克一郎 (KOMATSU, Katsuichiro) [JP/JP]; 〒145-0065 東京都大田区東雪谷2-15-9 Tokyo (JP).
 (26) 国際公開の言語: 日本語
 (30) 優先権データ:
 特願2001-11783 2001年1月19日 (19.01.2001) JP
 (74) 代理人: 小栗昌平, 外(OGURI, Shohei et al.); 〒107-6028 東京都港区赤坂一丁目12番32号 アーク森ビル28階栄光特許事務所 Tokyo (JP).
 (71) 出願人(米国を除く全ての指定国について): 三菱化学株式会社 (MITSUBISHI CHEMICAL CORPORATION) [JP/JP]; 〒100-0005 東京都千代田区丸の内二丁目5番2号 Tokyo (JP).
 (81) 指定国(国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK,

/統葉有]

(54) Title: METHOD OF CONSTRUCTING THREE DIMENSIONAL STRUCTURE OF PROTEIN INVOLVING INDUCED-FIT AND UTILIZATION THEREOF

(54) 発明の名称: 誘導適合を含めたタンパク質の立体構造構築方法およびその利用



(57) Abstract: It is intended to provide a method of accurately constructing the three-dimensional structure of an arbitrary protein. This method comprises selecting the normal vibration of a referential protein in the displacement direction, regarding the three-dimensional structure specified by the coordinates with displaced atomic coordinates as an induced-fit type referential protein, and thus constructing the three-dimensional structure of a target protein having an unknown three-dimensional structure. Thus, the three-dimensional structure of the target protein involving the normal mode of vibration (i.e., induced-fit) can be accurately constructed. This method makes it possible to accurately construct the structure of a protein more closely similar to the true structure, in particular, in the vicinity of a bond to a ligand. Therefore, this method is highly useful in, for example, designing molecules in the fields of medicine, agriculture, pharmacy, etc.

WO 02/057954 A1

BEST AVAILABLE COPY

/統葉有]



DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 國際調査報告書

(84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特

2 文字コード及び他の略語については、定期発行される各 PCT ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(57) 要約:

本発明は、任意のタンパク質の立体構造を精度良く構築する方法の提供を目的としてなされたものである。

本発明のタンパク質の立体構造構築方法は、参照タンパク質の基準振動を変位方向に選定し、原子座標を変位させた座標が規定する立体構造を誘導適合 (induced fit) 型参照タンパク質とし、立体構造が未知の目的タンパク質の立体構造を作成することよりなるものである。かくして、基準振動モード、即ち誘導適合 (induced fit) を含めた目的タンパク質の立体構造を精度良く構築することができる。本発明の方法は、より真に近いタンパク質の構造、特にリガンドと結合する近傍を精度良く構築しうる方法である。したがって、本発明の方法は医農薬分子の設計等に極めて有用である。

明細書

誘導適合を含めたタンパク質の立体構造構築方法およびその利用

技術分野

本発明は、誘導適合を含めたタンパク質の立体構造構築方法およびその利用に關し、さらに詳しく述べて、参考タンパク質の立体構造とその原子座標を変位させた複数の立体構造セットを参考タンパク質の立体構造として目的タンパク質の複数の立体構造セットを作成することによるタンパク質の立体構造構築方法、該立体構造セットを用いるタンパク質ーリガンド複合体の立体構造構築方法、およびタンパク質のリガンド結合部位の特定方法等に関する。

本発明の方法により提供される目的タンパク質の立体構造は、誘導適合(induced fit)を含めた立体構造であり、医農薬の分子設計に極めて有用である。

背景技術

立体構造が既知のタンパク質に関する情報を利用し、立体構造が未知の目的タンパク質とのアライメントを得て、このアライメント情報に基づいて目的タンパク質の立体構造を、コンピュータを用いて作成することが可能であり、この手法は、通常ホモロジーモデリング(homology modeling)と呼ばれている。ホモロジーモデリングにより構築される立体構造の精度は、近年目覚ましく向上しているが、未だ解決すべき問題点も多い。

この方法を用いて受容体タンパク質の立体構造を構築する場合、リガンドが結合する空間の確保が不可欠である。しかしながら、従来の立体構造構築法ではリガンドが存在する空間や結合部位に構築された立体構造の主鎖または側鎖がパッキングされ、その空間が塞がれてしまい、リガンドが受容体タンパク質と接触し、その結合部位に存在できない等の問題が生じていた。

また、タンパク質ーリガンド複合体の立体構造構築方法において、目的受容体タンパク質の立体構造が実験的に求まっていない場合、単にホモロジーモデリング(homology modeling)法により構築された受容体タンパク質の立体構造自身に

リガンドをドッキングさせ、分子力場計算や分子動力学計算で、それらを最適化することにより受容体タンパク質ーリガンド複合体の立体構造を得ていた。また、Multiple Copy Simultaneous Search (MCSS) 法を用いた研究においても、受容体タンパク質側の立体構造には基準振動モードは考慮されておらず、とくに分子の時間的にピコオーダーの振動を主とする長周期の熱搖らぎ(以下これを単に「熱搖らぎ」または「分子搖らぎ」と称することがある)は無視されていた。

更に、従来から、遠距離まで影響を及ぼす静電ポテンシャルによるタンパク質のリガンド結合部位を特定する方法や、類似化合物に基づいたタンパク質ーリガンド複合体の立体構造の構築が行われているが、いずれも信頼性が低く、類似化合物の無い場合には信頼性のあるタンパク質ーリガンド複合体の立体構造を導くことは困難であった。

発明の開示

本発明は、上記の状況を鑑みて、任意のタンパク質の立体構造を精度良く構築する方法、またタンパク質ーリガンド複合体の立体構造を精度良く構築する方法等の提供を目的としてなされたものである。

本発明者等は、上記課題を達成すべく鋭意検討した結果、参照タンパク質の原子座標を基準振動解析法から得られる固有ベクトル方向に変位した原子座標を参照して受容体タンパク質の立体構造を構築すれば、リガンドが存在する空間や結合部位に立体構造の主鎖または側鎖がパッキングされてその空間が塞がれることができず、受容体タンパク質の立体構造の精度を格段に向上させることができるを見出した。すなわち、基準振動モードに基づいて分子の熱搖らぎを考慮した複数の受容体タンパク質モデルが構築できることを見出した。

また、かくして構築された受容体タンパク質モデルにドッキングしたリガンドの立体構造を用いて、Multiple Copy Simultaneous Search (MCSS) 法の分子力学計算と分子動力学計算を適用して、分子の熱搖らぎを考慮した精度の高いタンパク質ーリガンド複合体の立体構造構築が可能であることを見出した。

更に、本発明者等は、タンパク質ーリガンド複合体には水溶液中の現象を考えると静電力よりも疎水相互作用の方が重要でないかという結論に達した。そこ

でタンパク質周囲および内部に溶媒を配置し、分子動力学による溶媒挙動（溶媒の拡散・集積）の解析からタンパク質に溶媒が集積する部位または溶媒が拡散しにくい部位が、リガンド結合部位と一致することを見出した。

本発明はこれらの知見に基づいて成し遂げられたものである。

即ち、本発明の方法により、(1) 参照タンパク質と目的タンパク質とのアライメントを導き出し、該アライメントおよび参照タンパク質の立体構造情報に基づいて目的タンパク質の立体構造を構築する方法において、参照タンパク質の立体構造とその原子座標を変位させた複数の立体構造を参照タンパク質の立体構造として目的タンパク質の複数の立体構造セットを作成することを特徴とする誘導適合を含めたタンパク質の立体構造構築方法が提供される。

この発明の好ましい態様により (2) 参照タンパク質の原子座標の変位が、基準振動解析法により行われることを特徴とする上記 (1) に記載の方法、(3) 立体構造の構築が、(i) アミノ酸中の $C\alpha$ 原子について参照タンパク質の立体構造から座標を取得し、目的関数を最小化するように $C\alpha$ 原子座標を最適化し、(ii) 最適化された $C\alpha$ の原子座標に主鎖の他の原子を付加して目的関数を最小化するよう主鎖の原子座標を最適化し、(iii) 最適化された主鎖の原子座標に側鎖の他の原子を付加し目的関数を最小化するよう最適化することにより行われることを特徴とする上記 (1) 又は (2) に記載の方法が提供される。

本発明の別の態様により、(4) (i) 上記 (1) ~ (3) のいずれかに記載の方法により得られる目的タンパク質の複数の立体構造とリガンドとのドッキング操作を行い、(ii) 目的タンパク質の 1 つの構造とリガンドとの構造の経験的分子エネルギー計算を、目的タンパク質の構造の数だけ行い、その際、(iii) 目的タンパク質側は、複数の構造それぞれのポテンシャルエネルギー勾配に応じて原子座標を動かし、(iv) リガンド側は、複数個算出されたポテンシャルエネルギー勾配を平均化した方向にリガンドの原子座標を動かして、(v) 目的タンパク質の複数の立体構造に基づくリガンドの立体構造を求める特徴とするタンパク質-リガンド複合体の立体構造構築方法が提供される。

この発明の好ましい態様により、(5) 経験的分子エネルギー計算において、目的タンパク質の初期 $C\alpha$ 原子座標の位置をオプション Harmonic 関数として加える

か、あるいは目的タンパク質の主鎖のねじれ角を拘束するポテンシャル関数を加えることを特徴とする上記（4）に記載の方法が提供される。

本発明の別の態様により、（6）（i）タンパク質の立体構造の周囲に低分子化合物を配置し、（ii）それらの周囲にさらに水分子を配置し、水溶媒中での経験的分子エネルギー計算を行って、タンパク質と低分子化合物との原子座標を得、（iii）得られた原子座標について、タンパク質の周囲および内部の、低分子化合物の挙動解析を行い、リガンドの結合部位を判定することを特徴とするタンパク質のリガンド結合部位の特定方法、および、（7）（i）タンパク質およびリガンドの立体構造の周囲に低分子化合物を配置し、（ii）それらの周囲にさらに水分子を配置し、水溶媒中での経験的分子エネルギー計算を行って、タンパク質と低分子化合物との原子座標を得、（iii）得られた原子座標について、タンパク質およびリガンドの周囲および内部の、低分子化合物の挙動解析を行い、タンパク質—リガンド複合体の結合部位を判定することを特徴とするタンパク質—リガンド複合体の結合部位の特定方法が提供される。

この発明の好ましい態様により、（8）低分子化合物の挙動解析が、低分子化合物を対象としたクラスター解析により行われ、得られたクラスターのサイズをリガンドの結合可能性部位の順位として結合部位を判定することを特徴とする上記（6）または（7）に記載の方法が提供される。

本発明の別の態様により、（9）上記（6）～（8）のいずれかに記載の方法により特定したタンパク質のリガンド結合部位にリガンドをドッキングし、経験的分子エネルギー計算によりタンパク質—リガンド複合体の立体構造を得ることを特徴とするタンパク質—リガンド複合体の立体構造構築方法が提供される。

本発明の別の態様により、（10）上記（1）～（5）および（9）のいずれかに記載の方法により得られるタンパク質の立体構造および／またはタンパク質—リガンド複合体の立体構造を規定する原子座標が記録されていることを特徴とするコンピュータ読みとり可能な記録媒体、または、該原子座標を含むことを特徴とするデータベースが提供される。

本発明の別の態様により、（11）上記（10）に記載の記録媒体またはデータベースから得られるタンパク質の立体構造を規定する原子座標を用いて、薬

物候補分子の立体構造との相互作用に基づいて、目的とする薬物分子を同定、検索、評価または設計することを特徴とする薬物分子設計方法が提供される。

図面の簡単な説明

第1図は、本発明の誘導適合を含めたタンパク質立体構造構築法の一例を示すフローチャートである。

第2図は、ステップ I-41 の $C\alpha$ 原子座標の構築方法を示す図である。アライメントの一致部分は参照タンパク質から取得し、無い部分は N, C両端それぞれ重なった 2 残基の重ね合わせの rmsd が最小のものをデータベースから取得する。

第3図は、ローカルスペースホモロジー (LSH) を示す図である。図中の T 残基に関する計算では、網をかけた（灰色の）残基が考慮される。図中下のアライメントにおける四角で囲った部分が考慮される残基ペアであり、* のマークがあるところの比率が LSH である。この場合 LSH は 56.2 % である。

第4図は、LSH と構造保存部位 (SCRs) にある比率との関係を示す図である。LSH は目的タンパク質と参照タンパク質との $C\alpha$ 原子の重ね合わせから計算され、SCRs にある比率は目的タンパク質の全残基数に対する Scrs 中の残基数である。

第5図は、本発明のタンパク質-リガンド複合体の立体構造構築法の一例を示すフローチャートである。

第6図は、本発明のリガンド結合部位の特定方法、該方法で特定された結合部位を用いるタンパク質-リガンド複合体の立体構造構築法の一例を示すフローチャートである。

第7図は、本発明の誘導適合を含めたタンパク質の立体構造構築方法の実施例の一例を示すフローチャートである。

第8図は、1F88(ロドプシン)を参照タンパク質として得られた QRHUB2(β 2アドレナリンレセプター)のアライメントを示す図である。図中、QRHUB2 及び 1F88 の右側の数字は、各々のタンパク質のアミノ酸配列においてアライメントの対象となったアミノ酸数である。また、上段の配列は QRHUB2(β 2アドレナリンレセプター)を示し、下段の配列は 1F88(ロドプシン)を示す。各タンパク質のアミノ酸配列は 1 文字記号で示す。

第9図は、最低固有値 4.47cm^{-1} の M' ($=26.4$)倍したゆらぎと換算した温度因子を示す図である。実線はPDB ID: 1F88のA鎖とB鎖平均の温度因子を換算したC α 原子のゆらぎであり、点線は基準振動解析法から得られた 4.47cm^{-1} の C α 原子位置ゆらぎを M' ($=26.4$)倍したものである。

第10図は、目的タンパク質と $\pm 2 \times M'$ ($\pm 2 \times 26.4$)倍した誘導適合 (induced fit) 型参照タンパク質から構築された誘導適合 (induced fit) 型目的タンパク質の立体構造の一部を示すディスプレイのプリントアウトの写真である。中央の構造が非誘導適合 (no induced fit) 型目的タンパク質である。

第11図は、MCSS 計算後のトリプシン-BPTI 複合体系のトリプシンの立体構造を示すディスプレイのプリントアウトの写真である。

第12図は、MCSS 計算前のトリプシン-BPTI 複合体の初期立体構造を示すディスプレイのプリントアウトの写真である。この図では、トリプシン-BPTI 複合体の活性部位に当たる、トリプシン側では His57、Asp102、オキシアニオンホール (Gly193-Asp194-Ser195) を、BPTI 側では Lys15 だけを抜き出して表示してある。図中、黒色で表示されている線がトリプシン-BPTI 複合体の X 線結晶解析の立体構造、灰色で表示されている線が組み立てられた複合体モデルの初期の立体構造である。

第13図は、MCSS 計算後のトリプシン-BPTI 複合体の立体構造を示すディスプレイのプリントアウトの写真である。この図では、トリプシン-BPTI 複合体の活性部位に当たる、トリプシン側では His57、Asp102、オキシアニオンホール (Gly193-Asp194-Ser195) を、BPTI 側では Lys15 だけを抜き出して表示してある。図中、黒色で表示されている線がトリプシン-BPTI 複合体の X 線結晶解析の立体構造であり、灰色で表示されている線が組み立てられた複合体モデルの精密化された立体構造である。

第14図は、トリプシン-BPTI 複合体の X 線結晶解析の立体構造座標を示すディスプレイのプリントアウトの写真である。

第15図は、トリプシン周囲のベンゼン分子の分布を示すディスプレイのプリントアウトの写真である。図中、黒線の六角形が一番大きなベンゼンクラスターである。

第16図は、BPTI周囲のベンゼン分子の分布を示すディスプレイのプリントアウトの写真である。図中、黒線の六角形が一番大きなベンゼンクラスターである。

発明を実施するための最良の形態

以下、本発明を更に詳細に説明する。本明細書において、幾つかの用語を使用するが、特に明記しない限り、次の意味を有する。

「目的タンパク質」とは、X線結晶解析やNMR解析等により完全な立体構造が決定されておらず、本発明において立体構造構築の対象となる任意のタンパク質を意味する。このタンパク質には、部分構造は解析されているが完全な立体構造が得られていないものも含まれる。本発明においては、立体構造が未知の受容体タンパク質、酵素等を目的タンパク質とするのが好ましい。ここで、X線結晶解析には、X線のみならず電子線および中性子線解析等も含まれる。

「受容体タンパク質」とは、細胞に存在し、外来性の物質あるいは物理的刺激を認識して、細胞に応答を誘起するタンパク質を意味する。この受容体タンパク質は、リガンドを特異的に結合する能力を有する。また、「リガンド」とは、タンパク質と特異的に結合する能力を有する物質を意味する。リガンドには、医農薬分子の様な低分子物質のみならず、抗体やタンパク質と相互作用をする特定のペプチドやタンパク質等の高分子物質も含まれる。

「参照タンパク質」とは、その立体構造の詳細がX線結晶解析やNMR解析等により既に決定されており、目的タンパク質の立体構造を規定する原子座標を構築するために参照するタンパク質を意味する。また「アライメント」とは、2種類以上のタンパク質についてアミノ酸配列の対応関係をつけることを意味する。

「原子座標」とは、三次元空間上で立体構造を記述するものである。それは空間上のある点を原点とする互いに垂直な三方向の相対的な距離であり、タンパク質中に存在する水素原子を除く原子一つあたりに3個の数字からなるベクトル量である。

「誘導適合 (induced fit)」とは、タンパク質の立体構造は柔軟であり、リガンド、例えば医農薬分子と結合すると、それとより良く結合するようにタンパク質の立体構造が変化することを意味する。「誘導適合 (induced fit)」を含めた立体

構造」とは、誘導適合により生じるタンパク質の立体構造変化を、例えば基準振動解析法で得られる固有ベクトルで表せると仮定し、誘導適合前の立体構造にこの固有ベクトルを加えて生成する立体構造を意味する。

「目的タンパク質ーリガンド複合体」とは、X線結晶解析やNMR解析等により複合体の完全な立体構造が解明されておらず、本発明において立体構造の構築対象となるタンパク質ーリガンド複合体を意味する。もちろんタンパク質としてX線結晶解析やNMR解析等により得られた立体構造を含むことは当然である。この複合体には、部分構造が解析されているが完全な立体構造が得られていないものも含まれる。タンパク質に結合したリガンド双方の複合体を意味する。

「Multiple Copy Simultaneous Search (MCSS) 法」とは、複数リガンドの立体構造を基にして目的タンパク質ーリガンド複合体の立体構造を、経験的分子エネルギー計算法、すなわち分子力学、分子動力学計算で受容体タンパク質の立体構造を求める方法である。本発明では、それとは逆に、複数のタンパク質の立体構造を1つのリガンドの立体構造を基に目的とするタンパク質ーリガンド複合体の立体構造を求める方法を意味する。

「経験的分子エネルギー計算」とは、分子力学計算と分子動力学計算を意味する。両者とも経験ポテンシャルを使った分子エネルギー計算である。

「MSAS (Maximum Solvent Accessibility of Sidechain)」とは、最大溶媒接触表面積のことであり、タンパク質を構成している各アミノ酸の側鎖の溶媒接触表面積と、そのアミノ酸がタンパク質を構成していない単独に存在する状態のときの側鎖の溶媒接触表面積との比を意味する。MSAS の詳細は、K. Akahane, Y. Nagano and H. Umeyama, Chem. Pharm. Bull., 1989, 37(1) 86-92 に記載されている。

後記 I ~ III の方法は、ホモロジーモデリングを行うことができる適当なコンピュータを用いて、後記方法を実行させる適当なプログラムを利用して実施することができる。

I. 誘導適合を含めた立体構造の構築方法

先ず、本発明の誘導適合を含めた立体構造の構築方法について説明する。

第1図は、本発明の誘導適合 (induced fit) を含めた立体構造構築方法の一例

を示すフローチャートである。

ステップ I-10において、目的タンパク質の配列を入力し、目的タンパク質の立体構造の構築に用いる参照タンパク質を選定し、参照タンパク質の立体構造から原子座標を取得し、目的関数を最小化するように原子座標を最適化する。ステップ I-20において、最適化した原子座標の基準振動解析法を行う。ステップ I-30において、固有ベクトル方向に参照タンパク質の原子座標を変位し、その構造を参照タンパク質に加え、参照タンパク質のセットを作成する。ステップ I-40において、適当なホモロジー・モデリング・プログラム、例えば FAMS によりアライメント情報や参照タンパク質セットの各立体構造情報から目的タンパク質の立体構造のセットを構築する。かくして、目的タンパク質の誘導適合 (induced fit) を含めた立体構造を精度良く構築することができる。以下、各ステップについて更に詳細に説明する。

ステップ I-10：参照タンパク質の初期座標の最適化

先ず、目的タンパク質の立体構造の構築において、目的タンパク質のアミノ酸配列を入力し、参照するタンパク質（参照タンパク質）を選定する。参照タンパク質の選定は、それ自体既知の通常用いられるアライメントソフトウェアを用いて行われる。この参照タンパク質の原子座標を、適当な立体構造データベースから取得する。この原子座標には、アミノ酸の骨格を作る窒素原子等に結合している水素原子はなく、ステップ I-20 の基準振動解析法の計算に水素原子が必要な場合は水素原子を発生させる。参照タンパク質の原子座標から構成される目的関数を用いて原子座標を最適化する。

ここで、用いられる目的タンパク質のアミノ酸配列としては、データベースに登録されているもの、配列が始めて解析されたもの等の如何なる由来の配列であってもよい。用いられるアミノ酸配列データベースとしては、例えば、"An Internet review: the complete neuroscientist scours the World Wide Web." Bloom FE, Science 1996; 274(5290): 1104-9 に詳細が記載されている GCRDb (The G-protein-coupled Receptor Database): <http://www.gcrdb.uthscsa.edu/>、 GPCRDB: <http://www.gpcr.org/7tm/>、 ExPASy: <http://www.expasy.ch/cgi-bin/sm-gpcr.pl>、 ORDB: <http://ycmi.med.yale.edu/senselab/ordb/>、 GeneBank:

ftp://ncbi.nlm.nih.gov/genbank/genomes/ 、 PIR: http://www-nbrf.georgetown.edu/pir/ (National Biomedical Research Foundation(NBRF))、 Swiss Plot: http://www.expasy.ch/sprot/sprot-top.html (Swiss Institute of Bioinformatics (SIB), European Bioinformatics Institute (EBI))、 TrEMBL (URL 及び管理者とともに Swiss Plot と同じ)、 TrEMBLNEW (URL 及び管理者とともに Swiss Plot と同じ)、 DAD: ftp://ftp.ddbj.nig.ac.jp(日本 DNA データバンク)等のデータベースに登録されているヒト (*H. sapiens*)、ショウジョウバエ (*D. melanogaster*)、線虫 (*C. elegans*)、酵母 (*S. cerevisiae*)、シロイヌナズナ (*A. thaliana*) 等を挙げることができる。これらのデータベースは単なる例示であり、タンパク質のアミノ酸配列が登録されているものであれば如何なるデータベースを用いることもできる。

また、参照タンパク質の原子座標の取得に用いられる立体構造データベースとしては、例えば PDB (Protein Data Bank): http://www.rcsb.org/pdb/、CCDC (Cambridge Crystallographic Data Centre: http://www.ccdc.cam.ac.uk/、SCOP (Structure Classification of Protein): http://scop.mrc-lmb.cam.ac.uk/scop、CATH: http://www.biochem.ucl.ac.uk/bsm/cath 等を挙げることができる。これらの立体構造データベースは、単独または組み合わせて用いることことができる。上記データベース中、SCOP および CATH は、ドメイン単位 (タンパク質の立体構造で、3 次構造の単位) に区切った立体構造データベースである。

アライメント用ソフトウェアとしては、例えば FASTA もしくは PSI-BLAST (Position-Specific Iterated BLAST) を使うのが好ましい。FASTA は目的配列と一致度の高い配列を立体構造データベースから探索し、最終的な目的配列と参照タンパク質との一致度を e 値として算出するプログラムである。FASTA の詳細は "Effective protein sequence comparison." Pearson WR, (1996) Methods Enzymol; 266:227-58 に記載されている。

PSI-BLAST はプロファイルアライメントを行うようにプログラムされている。PSI-BLAST の詳細は、"Matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices." Schaffer AA,

Wolf YI, Ponting CP, Koonin EV, Aravind L and Altschul SF, Bioinformatics 1999, 12, 1000-11 に記載されている。

参照タンパク質の原子座標の最適化を達成するための方法、座標系、目的関数等は特に制限されないが、例えば、最大傾斜法、共役勾配法、Newton-Raphson 法等で行うのが好ましい。最大傾斜法は、数値的に計算された目的関数の 1 次微分を利用し、原子座標の目的関数に対する最適化を行う。共役勾配法には、多くの方式があるが、Fletcher-Reeves 法 (Fletcher, R., and Reeves, C. M. (1964) Function Minimization by Conjugate Gradients. Comput J, 7: 149-154) が標準的に用いられており、目的関数の 1 次微分を利用し、目的関数が n 個の変数の厳密な二次関数である場合、多くとも n 回の繰り返しにより最適化に到達することが保証されている。Newton-Raphson 法は、1 次微分に加えて 2 次微分を利用し、初期構造が最適化構造に近い場合に効率が良い。これらの方法の詳細は、江口至洋「タンパク質工学の物理・化学的基礎（共立出版 1991）」とその中の文献に記載されている。

以下、上記の通り最適化した構造および座標を、それぞれ最適化構造および最適化座標として引用する。

ステップ I-20：最適化座標の基準振動解析法

上記ステップ I-10 で作成された最適化座標を用いて、その原子座標の変位を行う。原子座標の変位は、基準振動解析法を行い、各固有値の固有ベクトルを得ることにより行うのが好ましい。その際、最適化した自由度の一部を自由度とする座標系を用いても良い。この場合、一部の自由度に対しても最適化が達成されている。

ここで、「基準振動解析法」とは、ポテンシャルエネルギーを変位の 2 次関数として近似し、運動方程式を厳密に解き、最適化構造の周りの微小な振動を解析する方法を意味する。「固有値」とは、微小な振動の周期を意味する。「固有ベクトル」とは、振動の方向を意味する。

基準振動解析法の解くべき固有値方程式は、下記式(1)または(2)である。

$$\left(\sum_j T_{ij} U_{jk} \right) \omega_k^2 = \sum_j V_{ij} U_{ik} \quad (1) \text{ または } \mathbf{T} \mathbf{U} \boldsymbol{\Lambda} = \mathbf{V} \mathbf{U} \quad (2)$$

ただし、 $\Lambda_{ij} = \omega_i \delta_{ij}$, $\mathbf{U}^T \mathbf{T} \mathbf{U} = (\delta_{ij})$

ここで ω_k は固有値、 U_{ik} は固有ベクトルであり、 δ_{ij} はクロネッカーのデルタである。 T_{ij} と V_{ij} はそれぞれ運動エネルギー E_k とポテンシャルエネルギー V に関係し、下記式(3)および(4)の通りである。

$$E_k = \frac{1}{2} \sum_{i,j} T_{ij} \dot{q}_i \dot{q}_j \quad (3)$$

$$V = \frac{1}{2} \sum_{i,j} V_{ij} (q_i - q_i^0)(q_j - q_j^0) \quad (4)$$

ここで q_i は振動の自由度に対応した座標、 q_i^0 は最適化座標、 \dot{q}_i は q_i の時間による微分である。 A_{jk} は集団運動 Q_k と個々の原子運動 q_j を結ぶ係数であり、下記式(5)の通りである。

$$q_j = q_j^0 + \sum_k A_{jk} Q_k \quad (5)$$

ただし、基準振動 $Q_k = \alpha_k \cos(\omega_k t + \delta_k)$ である。

ここで、 α_k と δ_k は初期条件で定められる。

上記した基準振動解析法の詳細は、Wilson, E. B., Decius, J. C., and Cross, P. C. 1955. Molecular Vibrations. McGraw-Hill. に記載されている。

ステップ I-30：新規参照タンパク質の生成

上記ステップ I-20 で得られた固有値、固有ベクトルを用いて、ある温度・ある固有値での $C\alpha$ 原子の位置ゆらぎを計算する。固有値の数と等しい位置ゆらぎが得られる。参照タンパク質の $C\alpha$ 原子の温度因子を位置ゆらぎに換算し、各 $C\alpha$ 原子について基準振動解析法の位置ゆらぎとの比を計算し、平均の比を求める。この平均の比は、使用した固有値の数だけあり、この比を掛けたこの固有値に属す

る固有ベクトルを構造最適化前の参照タンパク質の原子座標に加え、この変位させた原子座標からなる立体構造、即ち、誘導適合（induced fit）を含めた立体構造を参照タンパク質の立体構造の1つとする。以下これを、誘導適合（induced fit）型参照タンパク質、立体構造、座標として引用する。

平均の比を2倍して同様に参照タンパク質の誘導適合（induced fit）型立体構造を作成する。固有ベクトルには順・逆の方向があり、固有ベクトルに-1を掛けた逆方向にも同様に変位させる。すなわち、誘導適合（induced fit）型には使用した固有値の数の4倍だけある。誘導適合（induced fit）型と非誘導適合（no induced fit）型参照タンパク質の立体構造を参照タンパク質立体構造セットとする。

ここで、温度因子と位置ゆらぎの関係は下記式(6)のとおりである。

$$D_i = \sqrt{\frac{3 \times B_i}{8 \times \pi^2}} \quad (6)$$

ここで、 B_i はPDBファイルから得られる原子の温度因子であり、 π は円周率、 D_i は位置ゆらぎに相当する。本発明では $C\alpha$ 原子に関してのみである。

基準振動法から得られる位置ゆらぎとPDBファイルの温度因子を換算した位置ゆらぎの比は下記式(7)のとおりである。

$$R_i^v = \frac{D_i}{F_i^v} \quad (7)$$

ここで F_i^v は基準振動解析法から得られる v 番目の固有値に対する i 番目の原子の位置ゆらぎである。本発明では、 $C\alpha$ 原子のみに対して行う。

比の平均は下記式(8)のとおりである。

$$M^v = \frac{\sum R_i^v}{N} \quad (8)$$

ここで N は原子数であり、和は原子に対して行う。 M^v は v 番目の固有値に対する平均の比である。本発明では、 $C\alpha$ 原子に対して行う。

誘導適合 (induced fit) 型参照タンパク質立体構造の原子座標は下記式(9)および(10)のとおりである。

$$C_{ik}^v = C_{ik}^0 \pm M^v \times V_{ik}^v : k = x, y, z \quad (9)$$

$$C_{ik}^v = C_{ik}^0 \pm 2 \times M^v \times V_{ik}^v : k = x, y, z \quad (10)$$

ここで C_{ik}^0 は参照タンパク質の原子座標、 V_{ik}^v は v 番目の固有値に属する固有ベクトルの成分をあらわす。

ステップ I-40：目的タンパク質のモデリング

上記ステップ I-30 で得られた参照タンパク質の立体構造セットを参照して、適当なホモロジー・モデリング・プログラム、例えば FAMS により目的タンパク質の立体構造セットを構築する。参照タンパク質の立体構造の数と同じ数の目的タンパク質の立体構造が構築される。即ち、使用した固有値の数の 4 倍だけある誘導適合 (induced fit) 型と非誘導適合 (no induced fit) 型目的タンパク質立体構造が構築され、これらを、目的タンパク質立体構造セット、すなわち誘導適合 (induced fit) を含めた立体構造とする。

次に、モデリング（立体構造の構築）手法の好適な一例として FAMS の各ステップについて説明する。なお、下記のステップ I-41～43において記載されている計算回数、定数、カットオフ値等は、本発明者が最も好ましいと考えているパラメータの一例を示すものであり、本発明の範囲を何ら限定するものではない。なお、FAMS の詳細は、Koji Ogata and Hideaki Umeyama, "An automatic homology modeling method consisting of database searches and simulated annealing" Journal of Molecular Graphics and Modeling 18, 258-272, 2000 に記載されている。

ステップ I-41：C_α 原子の初期座標の構築及び最適化

ステップ I-30 からの参照タンパク質セットおよびアライメント情報を受けて、参照タンパク質から挿入および欠損のあるアミノ酸残基についての情報をえる。アライメントにおいて連続して三残基以上のアミノ酸が対応しているギャップの無い領域を選び出し、その領域においては、これらの残基ペアにおいて、目的タ

ンパク質の $C\alpha$ 原子は参照タンパク質と同一のものを当てはめておく。 $C\alpha$ 原子が求められなかった場合には、予め作成してある断片のデータベースから座標を当てはめる（第2図参照）。

ここで、本明細書において $C\alpha$ 原子とは、各アミノ酸の骨格の中心となる炭素原子を意味する。 $C\beta$ 原子とは、 $C\alpha$ 原子の側鎖側に結合する炭素原子を意味し、 $C\gamma$ 原子とは、 $C\beta$ 原子の側鎖側に結合する炭素原子を意味する。また、 C 原子とは、カルボニル基の炭素原子を意味する。

ステップ I-41(1) : $C\alpha$ 原子のシミュレーティッドアニーリング法による構築

上記ステップ I-41 で作成された $C\alpha$ 原子はシミュレーティッドアニーリングのプロセスを用いて参考タンパク質の座標から構成される関数を用いて最適化される。この目的関数は下記式(11)のとおりである。

$$U_{C\alpha} = U_{len} + U_{ang} + U_{pos} + U_{vdw} \quad (11)$$

ここで U_{len} は、配列上隣の残基および Cys 残基のペアの $C\alpha$ 原子間の距離に関するもので下記式(12)のように設定される。

$$U_{len} = K_l \sum_i (D_{i,i+1} - 3.8)^2 + K_{ss} \sum_i (D_i^{ss} - 5.4)^2 \quad (12)$$

ここで $D_{i,i+1}$ は残基 i と残基 $i+1$ の $C\alpha$ 間距離である。 D_i^{ss} はジスルフィド結合を形成する Cys 残基のペア同士の距離である。 K_l と K_{ss} は定数でありそれぞれ 2 および 5 と設定される。

U_{ang} は $C\alpha$ 原子の結合角の関数であり下記式(13)のとおりである。

$$U_{ang} = K_a \sum_i (\theta_i - \theta_0)^2 \quad (13)$$

ここで θ_i (rad) は $i, i+1, i+2$ 番目の残基 $C\alpha$ 原子の角度である。 θ_0 は PDB の X 線構造から $(100/180) \cdot \pi$ (rad) と設定される。 K_a は定数であり 1 とする。

U_{pos} は、 $C\alpha$ 原子の位置に関する関数であり、下記式(14)のとおりである。

$$U_{pos} = K_{pos} \sum_i \frac{1}{M_i} \| x_i - \langle w_i x_i \rangle \|^2 \quad (14)$$

ここで $\| \cdot \|$ が意味する所はノルムであり、 M_i は構造を基にしたアライメント上で構造的に等価な位置にある $C\alpha$ 原子間の平均距離である。残基 i について M_i の値が求められないとき、 M_i の値は 10 と設定される。ここでは、 $C\alpha$ 原子の平均座標であり下記式(15)のとおりである。

$$\langle w_i x_i \rangle = \frac{1}{W} \sum_j w_i^j x_i^j \quad (15)$$

ここで x_i^j は j 番目の参照タンパク質の i 番目の残基の $C\alpha$ の原子座標である。 w_i^j は、 j 番目の参照タンパク質の i 番目の残基の重みである。この重みは目的タンパク質の大体の形を決定するため重要なパラメータであるが、これはローカルスペースホモロジー (LSH) と呼ばれる着目部位の 12 Å 以内の空間的近傍の局所的な値によって決定している (第 3 図参照)。LSH と構造がよく保存されている部位 (SCRs : Structural Conserved Regions) に存在する残基のペアの比率との相関は第 4 図に示されているように非常に高い。これは、高い LSH 値を持つときは統計的に $C\alpha$ 原子の位置が参照タンパク質構造と比べて 1.0 Å 以内にあることを意味する。

U_{vdw} は下記式(16)のとおりである。

$$U_{vdw} = K_{vdw} \sum_{i,j (>i+2)} \left\{ \left(\frac{3.8}{D_{i,j}} \right)^{12} - \left(\frac{3.8}{D_{i,j}} \right)^6 \right\} \quad (16)$$

ここで K_{vdw} は $0.01 (D_{i,j} < 3.2 \text{ Å})$ と $0.001 (D_{i,j} > 3.2 \text{ Å})$ と設定され 6 Å をカットオフ値とした。

$C\alpha$ 原子は式(11)に従って、シミュレーティッドアニーリング法を用いて最適化される。この最適化の段階で $C\alpha$ 原子の摂動は 1.0 Å 以内になるように設定する。またこのアニーリングの段階は全ての $C\alpha$ 原子について、100 回づつ計算される。

そして、温度に相当するパラメータは、25から0.5回ごとに0.01減らし、そのパラメータは以後一定とした。

この大きな2つの段階、構造情報の取得とC_α原子の構築は10回繰り返され、最小の目的関数値をもつC_α原子の座標が最適解として算出される。

ステップ I-42：主鎖原子座標の構築及び最適化

ステップ I-41(1)のC_αの原子座標に主鎖の他の原子を付加し、シミュレーティッドアニーリング法によって目的関数を最小化するようとする。まず、C_α原子の立体的な重ねあわせを行い、C_αの原子間距離が2.5Å以下の残基が取り上げられる。C_αを除く主鎖の原子座標はC_α原子間距離が最小になるように参照タンパク質の座標から取得しモデル構造とする。

参照タンパク質の中に相当する残基が無い場合、主鎖の原子座標はデータベース中の相当する4残基のタンパク質断片から作成される。この過程の中で、残基*i*の主鎖原子は*i*-1番目から*i*+2番目までのC_α原子間の最小のrmsd値を持つ残基から選ばれる。その際N末端の残基では、C_α原子座標の重ね合わせ範囲が*i*番目から*i*+3番目までとなり、C末端の残基およびそのひとつ前の残基では同様に*i*-3番目から*i*番目までおよび*i*-2番目から*i*+1番目までとなる。

主鎖原子の目的関数を元にシミュレーティッドアニーリング法によって最適化される。

目的関数は下記式(17)のとおりである。

$$U_{main} = U_{bond} + U_{ang} + U_{non-bond} + U_{ss} + U_{pos} + U_{tor} + U_{chi} + U_{hydr}$$

(17)

U_{bond}は下記式(18)のとおりである。

$$U_{bond} = K_b \sum_i (b_i - b_i^0)^2 \quad (18)$$

ここでb_i⁰は、標準の結合長でありそれぞれの化学結合の種類によって異なる。

K_bは定数であり225とする。

U_{ang}は結合角の関数で、下記式(19)のとおりである。

$$U_{ang} = K_a \sum_i (\theta_i - \theta_i^0)^2 \quad (19)$$

ここで θ_i は i 番目の結合角であり、化学結合の種類によって異なる。 K_a は定数で 45 と設定される。

$U_{non-bond}$ は非結合の相互作用の関数で、下記式(20)のとおりである。

$$U_{non-bond} = K_{non} \sum_{i,j} \varepsilon_{i,j} \left\{ \left(\frac{r_{i,j}^*}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^*}{r_{i,j}} \right)^6 \right\} \quad (20)$$

ここで $\varepsilon_{i,j}$ と $r_{i,j}^*$ は定数で原子の種類によって異なる。

K_{non} は定数で 0.25 とし、カットオフは 8Å とする。

U_{ss} は Cys 残基が生成するジスルフィド結合の関数で、下記式(21)のとおりである。

$$U_{ss} = \sum_i \left\{ K_{C\alpha}^{ss} (D_i^{C\alpha} - 5.4)^2 + K_{C\beta}^{ss} (D_i^{C\beta} - 3.8)^2 \right\} \quad (21)$$

ここで $K_{C\alpha}^{ss}$ および $K_{C\beta}^{ss}$ は定数であり 7.5 である。

U_{pos} は原子の位置に関する関数で、下記式(22)のとおりである。

$$U_{pos} = K_{pos} \sum_i \frac{1}{M_i} \| \mathbf{x}_i - \langle w_i \mathbf{x}_i \rangle \|^2 \quad (22)$$

ここで $\langle w_i \mathbf{x}_i \rangle$ は下記式(23)のように与えられる。

$$\langle w_i \mathbf{x}_i \rangle = \frac{1}{W} \sum_j w_i^j \mathbf{x}_i^j \quad (23)$$

式(22)の $\langle w_i \mathbf{x}_i \rangle$ は、目的タンパク質および参照タンパク質の間の構造の重ねあわせから求める。

K_{pos} は定数であり 0.3 である。

U_{tor} は主鎖のねじれ角のものであり、下記式(24)のとおりである。

$$U_{tor} = K_t \sum_i \sqrt{(\phi_i - \phi_i^0)^2 + (\psi_i - \psi_i^0)^2} + K_\omega \sum_i (\omega_i - \omega_i^0)^2 \quad (24)$$

ここで ϕ_i^0 と ψ_i^0 は Ramachandran マップ上での最も近いねじれ角の ϕ_i および ψ_i とする。また ω_i^0 は 0 として cis-Pro 残基の場合のみ π (radian) とする。 K_t および K_ω は定数であり、それぞれ 10 および 50 とする。

U_{chi} は $C\alpha$ のキラリティーに関するものであり、下記式(25)のとおりである。

$$U_{chi} = K_{chi} \sum_i \left(\tau_i + \frac{2}{3}\pi \right)^2 \quad (25)$$

ここで τ_i は $N-C\alpha-C\beta-C$ で定められるねじれ角であり K_{chi} は 50 とする。

U_{hydr} はホモロガスなタンパク質中で保存された主鎖の水素結合に関するもので、下記式(26)のとおり定められる。

$$U_{hydr} = K_{hydr} \sum_{i,j} (D_{i,j}^{N-O} - 2.9)^2 \quad (26)$$

水素結合は、N 原子と O 原子の距離が 2.9 ± 0.5 Å にあるときに設定される。

複数の参照タンパク質中で水素結合があるか否かを判定するときは、75%以上の参照タンパク質が存在すると認めた場合に水素結合ありと判定する。 K_{hydr} は定数であり 0.6 である。

次に $C\beta$ を含む主鎖原子の最適化がシミュレーティッドアニーリングによって行われる。このアニーリングの過程で主鎖と $C\beta$ の原子の振動が初期の位置に対して 1.0 Å 以内になるようにする。このアニーリングの段階は主鎖と $C\beta$ の原子に対して 200 回行われる。温度に相当するパラメータは 50 もしくは 25 から始まり一回毎に 0.5 倍にしてゆき 0.01 になるまで続け、その後一定値とする。

主鎖の立体配置を幅広くサンプリングするために、本発明の方法では、好ましくは上記の方法を 6 回行い、最小の目的関数值を持つ主鎖の原子座標を最適解とする。そして、温度に相当するパラメータは、はじめの 2 回は 50 からスタートして 3 回目から 25 からスタートすることとする。

ステップ I-43：側鎖原子座標の構築及び最適化

側鎖の構築は、大きく2段階に分かれており、「構造保存部位の側鎖構築」(ステップ I-43(1)) と「全体の側鎖構築」(ステップ I-43(2)) に分けられる。

ステップ I-43(1)：構造保存部位の側鎖構築

算出された主鎖原子に対して、以前の研究における方法を用いてホモガスなタンパク質から側鎖のねじれ角を得る。この方法の詳細は、“The role of played by environmental residues in side-chain torsional angles within homologous families of proteins: A new method of side chain modeling.” Ogata K and Umeyama H, Prot. Struct. Funct. Genet. 1998, 31, 255-369 に記載されている。

この方法の中でホモガスなタンパク質中で保存されている側鎖の割合を算出し、この情報を元にして側鎖のモデリングを行う。側鎖の保存された部位の側鎖の原子座標は固定した主鎖原子に対して置かれる。例えば、ホモガスなタンパク質中でアルギニン残基の χ^1 角が保存されていれば、 $C\gamma$ 原子の座標を置くことができ、Phe 残基で χ^1 と χ^2 角が保存されていれば、全ての側鎖原子を置くことができる。式(17)を用いたシミュレーティッドアニーリングの最適化の過程は、主鎖と $C\beta$ の原子のみ行われて、原子の振動は 1.0 Å 以内となるようにした。この主鎖と $C\beta$ の原子のアニーリングの段階は 200 回行われる。そして、温度に相当するパラメータは 25 からスタートして一回毎に 0.5 倍にしてゆき 0.01 になるまで小さくなるようにする。式(17)の中の $U_{\text{non-bond}}$ は主鎖原子と部分的に作成された側鎖原子について行われる。そのとき側鎖原子の座標は最適化の過程を通じて保存されるようとする。

構造の情報である M_i と水素結合の N-O のペアは最適化の過程で用いられる。主鎖原子の配置を得るために、上記プロセスを 3 回繰り返し、目的関数の最小の主鎖原子の座標を算出構造とする。

ステップ I-43(2)：全体の側鎖の構築

側鎖の構築は固定した主鎖および $C\beta$ 原子のもとで行う。これは上記した Ogata K and Umeyama H, Prot. Struct. Funct. Genet. 1998, 31, 255-369 に開示されている研究成果をもって行われ、それを用いることにより短時間で正確なモデルを与えることができる。次に主鎖構造は低温におけるモンテカルロ法によって最

適化され、温度は 0.001 に設定され式(17)の目的関数 $U_{\text{non-bond}}$ を用い、全ての主鎖と側鎖の原子で計算される。そして、N、C α 、C、C β 原子の最適化の過程で側鎖のねじれ角を最適化された状態を保つように側鎖の座標を再配置する。原子の振動は 0.5 Å 以内とする。次に側鎖は削除され、上記の側鎖構築が繰り返される。このプロセスは 2.4 Å の原子同士のぶつかり合いがなくなり、且つ N-C α -C β -C のねじれ角が -120 ± 15° の範囲に収まるまで繰り返される。

ステップ I-44：最終構造の構築

かくして、任意の目的タンパク質の非誘導適合 (no induced fit) 型と誘導適合 (induced fit) 型の立体構造を規定する原子座標を得ることができる。

II. タンパク質-リガンド複合体の立体構造構築方法

次に、本発明の別の態様であるタンパク質-リガンド複合体の立体構造構築方法について、図面を参照して説明する。第 5 図は、目的とするタンパク質-リガンド複合体の立体構造構築方法、即ち誘導適合 (induced fit) を含んだ複合体の立体構造構築方法の一例を示すフローチャートである。

まず、ステップ II-10において、目的タンパク質のモデリングされた原子座標を得る。最適化された参考タンパク質の基準振動解析法を行なうことにより、基準振動モードが得られる。そして固有ベクトル方向に主に実験で得られた目的タンパク質の原子座標を変位し、複数の参考タンパク質のセットを作成する。それらの座標を参照して目的タンパク質の立体構造をホモジーモデリング (homology modeling) により構築する。

ステップ II-20 で、得られた目的タンパク質の立体構造に対してリガンドをドッキングさせる操作を行なう。ステップ II-30 において、目的タンパク質のセットにドッキングしたリガンドに基づき MCSS 法による経験的分子エネルギー計算を行ない、目的とするタンパク質-リガンド複合体の立体構造をシミュレートする。かくして得られる目的タンパク質-リガンド複合体の立体構造は、目的タンパク質の誘導適合 (induced fit)、すなわち周期的熱運動 (分子揺らぎ) を含めた立体構造であり、医農薬の精度の高い分子設計に用いることができる。

以下、各ステップについて更に詳しく説明する。

ステップ II-10：目的タンパク質のモデリング

目的タンパク質のモデリングは、次の3つのステップ II-11：参照タンパク質の初期座標の最適化、II-12：最適化座標の基準振動解析、II-13：目的タンパク質のモデリングに分けられる。このステップは、前記 I-10～I-44 と同様に行われる。かくして、基準振動解析法の振動モードに基づく立体構造、すなわち誘導適合(induced fit)を含んだ目的タンパク質の立体構造が構築できる。

ステップ II-20：目的タンパク質へのリガンドのドッキング

基準振動モードを考慮した目的タンパク質の複数の立体構造モデルに対してリガンドのドッキングを行なう。目的タンパク質のリガンド結合サイトと考えられる位置にドッキングさせる。このステップは、PDB 形式のファイルが入出力できる市販のソフトウェア、例えば BIOCES(NEC 社製)、Cerius2(Accelrys 社製)、SYBYL(TRIPOS 社製)、HyperChem(Hypercube 社製)等を用いて行なう。一般的にはドッキングはステレオ表示が可能なディスプレイ上でリガンドを回転、並進して行なう。また簡易的エネルギー計算手法を含めたドッキングを行なってもよい。

用いるリガンドの結合部位は、特に限定されず、既に判明している結合部位、新たに特定した結合部位のいずれも用いることができる。リガンドの結合部位が未知のタンパク質については、後記 III で述べる方法により、その部位を特定することもできる。

ステップ II-30：目的タンパク質—リガンド複合体の立体構造の最適化

ステップ II-20 で得られたタンパク質—リガンド複合体構造モデルについて、目的タンパク質の1つの構造とリガンドとの構造の経験的分子エネルギー計算を、目的タンパク質の構造の数だけ行い、その際、目的タンパク質側は、複数の構造それぞれのポテンシャルエネルギー勾配に応じて原子座標を動かし、リガンド側は、複数個算出されたポテンシャルエネルギー勾配を平均化した方向にリガンドの原子座標を動かして、目的タンパク質の複数の立体構造に基づくリガンドの構造を求める。

このステップ II-30 は、例えば Multiple Copy Simultaneous Search (MCSS) 法により行われ、リガンドにより複数の複合体構造が経験的分子エネルギー計算(分子力場法)により同時に最適化され、それらの原子座標は経験的分子エネルギー

計算（分子動力学法）により、構造が、例えば温度 300° K で 10ps 間緩和され、さらにその原子座標は分子力場法により最適化されることにより行われる。もちろん温度、時間は計算している対象系によって変わることはある。

MCSS 法は、複数のリガンドを用いて受容体タンパク質とリガンド双方の立体構造を最適化する手法として A. Miranker and M. Karplus (Proteins, 1991, 11, 29-34) により提案されている。手法としては、個々のリガンドとタンパク質の経験的分子エネルギー計算を同時に行ない、受容体タンパク質のグラジエントについては平均をとるため、受容体タンパク質側は 1 つの立体構造として動く。

これに対して、本発明の方法では、タンパク質側は複数の分子構造、リガンド側は 1 つの分子構造を用いて、複数のタンパク質構造に基づくリガンドの構造を求めるものである。この時の経験的分子エネルギー計算において、タンパク質 1 構造とリガンド 1 構造の計算を、タンパク質構造の数だけ行い、リガンド側は、複数個算出されたポテンシャルエネルギー勾配を平均化した方向にリガンドの原子座標を動かす。一方、目的タンパク質側は、複数の構造それぞれのポテンシャルエネルギー勾配に応じて原子座標を動かし、目的タンパク質の複数の立体構造に基づくリガンドの構造が求められる。

上記の経験的分子エネルギー計算の方法は、特に限定されずそれ自体既知の方法で行えば良いが、発明者らが開発した apricot プログラム (Yoneda, S., and Umeyama, H., J Chem Phys 1992; 97: 6730-6736) を改良した apricot-MCSS プログラムを用いるのが好ましい。経験的ポテンシャル関数としては AMBER タイプのポテンシャル関数 (S. J. Weiner, P. A. Kollman, D. A. Case, U. Chandra Singh, C. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner, J. Am. Chem. Soc., 1984, 106, 765-784) を、パラメータとしては parm89a Rev A を用いるのが好ましい。もちろん他の経験ポテンシャルの使用も可能である。

分子動力学計算では、通常のエネルギー項の他に $C\alpha$ 原子位置に対する拘束ポテンシャルを、例えば下記式(27)のように Harmonic 関数として加えることにより、目的タンパク質の初期立体構造が大きく壊れないようにするのが好ましい。これは計算の近似の粗さを補う意味で大切であるが、拘束ポテンシャルの範囲を主鎖全体に広げたりしてもよく、これに限定されるものではない。

$$U_{xyz} = \sum K_{xyz} (X - X_0)^2 \quad (27)$$

ここで U_{xyz} は目的タンパク質における $C\alpha$ 原子位置に掛ける拘束のポテンシャルエネルギーで、 $C\alpha$ のオリジナル座標値が x_0 、更新された座標値が x 、 K_{xyz} が原子をどの程度拘束させるかのパラメータである。ここでは K_{xyz} として $10.0 \text{ kcal/mol}/\text{\AA}^2$ を用いたが、一例であるので、式の形を含めて本発明の範囲を限定するものではない。

また、 $C\alpha$ 原子の X、Y、Z 座標に対する拘束ポテンシャルの代りに、式(24)に示す目的タンパク質の主鎖のねじれ角に対する拘束を用いて、すなわち経験的分子エネルギー計算において目的タンパク質の主鎖のねじれ角を拘束するポテンシャル関数を加えることにより、初期立体構造が大きく壊れないようにもよい。

かくして、目的タンパク質として誘導適合 (induced fit) 型の立体構造モデルを使えば、分子揺らぎを考慮した目的タンパク質—リガンド複合体の原子座標を得ることができる。

また、リガンド分子がタンパク質の場合には、上記と同様の方法で、リガンドの基準振動モードを含む複数の立体構造とタンパク質の単一立体構造からリガンド側の誘導適合 (induced fit) を考慮したリガンド—タンパク質複合体の立体構造の構築も可能である。

III. タンパク質のリガンド結合部位の特定方法

次に、本発明の別の態様である、タンパク質のリガンド結合部位の特定方法について説明する。第6図は、タンパク質のリガンド結合部位の特定方法と、得られた結合部位にリガンドを結合させて、タンパク質—リガンド複合体の立体構造を構築する方法の一例を示すフローチャートである。

ステップ III-10 で、タンパク質とリガンドの結合部位の特定（予測）を行う。このステップにおいて、タンパク質および／またはリガンドの周囲および内部、例えば疎水性表面に、低分子化合物、例えば非極性溶媒を発生させ、さらにそれら周囲に多数の水分子を追加して見かけ上水溶液中の分子動力学計算を行う。そ

これらの結果に基づき、タンパク質および／またはリガンド表面の低分子化合物、例えば非極性溶媒の挙動から、タンパク質とリガンドの結合部位を検索する。ステップ III-20 ではステップ III-10 で得られたタンパク質とリガンドの結合推定部位を参考にして、それらをドッキングさせ、タンパク質—リガンド複合体の立体構造の初期原子座標を求める。そしてステップ III-30 ではステップ III-20 で得られたタンパク質—リガンド複合体の初期立体構造の周囲に水分子を発生させ、見かけ上水溶媒中の分子力学と分子動力学法を用いてタンパク質—リガンド複合体の立体構造の精密化を行う。

以下、各ステップについて更に詳しく説明する。

ステップ III-10：タンパク質のリガンド結合部位の特定

タンパク質とリガンドの結合部位の特定は、次の 3 つのステップ、III-11：タンパク質周囲および／またはリガンド周囲への低分子化合物の発生、III-12：タンパク質および／またはリガンドの水溶媒中での経験的分子エネルギー計算（分子力学、分子動力学計算）による低分子化合物（例えば非極性溶媒等）の挙動検索、III-13：低分子化合物（例えば非極性溶媒等）の挙動から、タンパク質へのリガンド結合部位および／またはリガンドのタンパク質への結合部位の判定に分けられる。

ステップ III-11：タンパク質周囲および／またはリガンド周囲への低分子化合物の発生

先ず、タンパク質および／またはリガンドの周囲に水分子を発生させたのち、タンパク質周囲、リガンド周囲、ならびに低分子化合物が入り込める内部周囲にある水分子を低分子化合物で置換する。その場合、これらの置換はそれら周囲全体にわたり低分子化合物を配置してもよいし、疎水性や水素結合能を有するアミノ酸や官能基の周りにだけ低分子化合物を配置してもよい。ここで、リガンドがペプチドやタンパク質等の高分子物質である場合には、リガンド周囲へも低分子化合物を発生させ、タンパク質の場合と同様に経験的分子エネルギー計算による低分子化合物の挙動解析を行う。リガンドが医農薬分子等の分子量が小さい物質である場合は、どの部分が疎水性領域か等の判別できるので、通常、結合部位の特定の必要性は無い。しかし、リガンドが高分子物質である場合は、リガンド側

の結合部位もタンパク質側の結合部位と同様に解析し、複合体の結合部位を特定することが必要である。

低分子化合物としては、例えば、エタン、シクロヘキサン、ベンゼン等の非極性溶媒、N-メチルアセタミド、ベンズアミド等の水素結合能性溶媒、あるいは医農薬化合物でもよく特に限定されない。だがそれら配向の任意性を考えると、対象性を有する化合物が好ましい。非極性溶媒を用いると疎水性部分を有するタンパク質やリガンドの結合部位を特定することができる。また水素結合能性溶媒である酸アミド基を有する化合物を用いると、酸アミド基と水素結合しうる部分、すなわち β シート構造の露出部分やオキシアニオンホールを含むリガンドの結合部位を特定することができる。更に医農薬分子を用いると、医農薬分子が特異的に結合しうる部分を特定することができる。

具体的には、例えばベンゼン等の非極性溶媒をタンパク質の周囲に配置させる場合は、タンパク質の中でMSASの値が30%以上のアミノ酸残基により形成される3.5Å以内の表面にある水分子を非極性溶媒（ベンゼン）で置換すれば良い。また非極性溶媒（ベンゼン）同士が1.5Å以内になるような場合には水分子の非極性溶媒への置換は行わなくて良い。非極性溶媒に置換されなかつた水分子は一回すべて消去する。上記した水分子の非極性溶媒への置換基準は、ベンゼンを用いた場合の一例であり、本発明の範囲を限定するものではない。

ステップ III-12: タンパク質および／またはリガンドの水溶媒中での経験的分子エネルギー計算による低分子化合物の挙動検索

上記ステップ III-11で作成されたタンパク質（および／またはリガンド）と低分子化合物の原子座標を用いて、それら周囲に周期境界条件で水分子を発生させたのち、経験的分子エネルギー計算である分子力学計算で立体構造を最適化し、続いて分子動力学計算を行う。分子動力学計算が終了したのち、水分子を除去してタンパク質（および／またはリガンド）と低分子化合物との原子座標を得る。例えば、低分子化合物として非極性溶媒（ベンゼン）を配置した場合は、温度300K、10~20ps程度の分子動力学計算を行えば良い。これにより、タンパク質の周囲や内部への低分子化合物の拡散や集積が起こる。この拡散や集積の状態、即ち低分子化合物の挙動を、後記ステップ III-13の方法で解析することにより、タン

パク質側のリガンド結合部位、リガンド側のタンパク質結合部位を特定することができます。

上記の経験的分子エネルギー計算の方法は、特に限定されないが、本発明者らが開発した apricot プログラムを用いるのが好ましい。経験的ポテンシャル関数としては AMBER タイプのポテンシャル関数を用いるのが好ましい。もちろん他の経験ポテンシャルの使用も可能である。

ステップ III-13：低分子化合物の挙動からのリガンド結合部位の判定

上記ステップ III-12 で求まったタンパク質周囲および／またはリガンド周囲の低分子化合物、例えば非極性溶媒の分布について、これを対象としたクラスター解析を行い、得られたクラスターの大きさからリガンドがタンパク質にドッキングしやすい部位を判定する。

ここで、クラスター解析とは、多次元空間において与えられたデータ集合を個体間の類似度(あるいは相違度)によってクラスター(塊)化する多変量解析法である。ここでは 3 次元空間における非極性溶媒の重心(ベンゼンでは 6 炭素原子の座標平均)間同士のユークリッド距離を計算し、閾値以内の距離の非極性溶媒があれば、距離が短い非極性溶媒同士からクラスター化していく。そのときクラスター化された非極性溶媒の集合についても、通常のクラスター解析と異なり、クラスターの重心からの距離ではなく、その中で最短距離の非極性溶媒同士が閾値以内であるかどうかを調べることにより、それらをクラスター化するか否かを判定する。非極性溶媒のベンゼンの場合、閾値については 6 Å を用いたが、その値は單なる例示であり本発明の範囲を限定するものではない。

例えば非極性溶媒(ベンゼン)を用いた場合、それらはいくつかのクラスターに分類されるが、大きなクラスターほどリガンドやタンパク質へのドッキング部位である可能性が高いと考えられる。クラスター化された非極性溶媒群はその形状を橍円球で表現できるが、座標の固有値問題を解くことにより、クラスターの長短方向が求まる。タンパク質側とリガンド側双方のクラスター同士を橍円球の長短方向を参考にしてドッキングし、タンパク質-リガンド複合体のモデルをいくつか作成する。もちろんタンパク質とリガンドが重なる配置になる複合体構造は自動的に取り除く。ドッキングされたモデルはステップ II-20 で記述したソフ

トウエアでタンパク質とリガンド配置の微調整を行う。

ステップ III-20：タンパク質へのリガンドのドッキング

上記ステップ III-13 で得られた低分子化合物、例えば非極性溶媒（ベンゼン）のクラスタリングで大きなクラスターとなったサイト同士をドッキングし、タンパク質-リガンド複合体構造の初期データとする。この際、低分子化合物、例えば非極性溶媒（ベンゼン）データはドッキングに際して除かれる。

本ステップは PDB 形式のファイルが入出力できる市販のソフトウェアを用いて行なうことができる。一般的にはドッキングはステレオ表示が可能なディスプレイ上でリガンドの回転、並進等により行なわれる。また簡易的エネルギー計算手法を含めたドッキングを行なってもよい。

ステップ III-30：タンパク質-リガンド複合体の立体構造の構築

上記ステップ III-20 で得られたタンパク質-リガンド複合体の初期原子座標データは、それら周囲に周期境界条件で水分子を発生させたのち、分子力学計算で初期立体構造を最適化し、続いて分子動力学計算を行い、そして最終ステップの座標軌跡から水分子を取り除くことによりタンパク質-リガンド複合体の立体構造が得られる。

分子動力学計算の方法は、特に限定されず、例えば、温度 300° K、10 から 20ps 程度で行なえばよい。用いるプログラムも特に限定されないが、発明者らが開発した apricot で、経験力場も AMBER タイプを用いるのが好ましい。しかし使用プログラム、力場とも単なる例示であり、本発明の範囲を限定するものではない。

かくして、タンパク質-リガンド複合体の生成過程が水溶液中であることを考慮して、水溶媒中での低分子化合物、例えば非極性溶媒の集積、拡散を利用して、タンパク質とリガンドの疎水性表面を見い出し、それら同士をドッキングするという方法でこれまでより精密なタンパク質-リガンド複合体の原子座標を得ることができる。

IV. タンパク質の立体構造を規定する原子座標が記録されている記録媒体、データベース

上記方法で得られたタンパク質の立体構造またはタンパク質-リガンド複合

体の立体構造を規定する原子座標を、コンピュータが利用可能な所定の形式で適當な記録媒体に格納することにより、目的タンパク質の立体構造データベースが構築できる。本発明のデータベースは、好ましくは、上記原子座標とともに参照タンパク質と目的タンパク質のアライメント情報を含んでいても良い。また、データベースには、所望によりコード番号、参照タンパク質の参照領域の情報、目的タンパク質の情報、C_α原子間距離等が含まれる。

本発明においてデータベースとは、上記原子座標を適當な記録媒体に書き込み、所定のプログラムに従って検索を行うコンピュータシステムをも意味する。ここで適當な記録媒体としては、例えば、フロッピーディスク、ハードディスク、磁気テープ等の磁気媒体；CD-ROM、MO、CD-R、CD-RW等の光ディスク、半導体メモリ等を挙げることができる。

V. 薬物の分子設計方法

医農薬等の薬物分子設計を行うことができる適當なプログラムが動作するコンピュータで、上記方法で得られた薬物分子の標的となるタンパク質（以下これを「標的タンパク質」と称することがある）の構造座標の全て若しくは一部、又はそれらが記録されたデータベース若しくは記録媒体の構造座標の全て若しくは一部を使用して、標的タンパク質と相互作用をする薬物分子（拮抗薬または作動薬）を同定、検索、評価又は設計等を行うことができる。

薬物分子の同定、検索、評価又は設計は、本発明の方法で得られた立体構造座標と薬物分子の立体構造座標との相互作用の有無やその程度に基づいて行われる。本明細書において、薬物分子の同定、検索、評価又は設計等を、単に薬物の分子設計ということがある。

タンパク質の立体構造座標と薬物候補分子の立体構造座標との相互作用に基づいて分子設計を行う際に用いられるコンピュータとしては、適當なプログラムが動作するように調整されているコンピュータであれば特に制限はない。また、コンピュータの記憶媒体にも特に制限はない。分子設計に用いるプログラムは、例えばアクセルリス(Accelrys)社製のコンピュータ・プログラム Insight II等を挙げることができる。特に、この目的のために特別に作成された Ludi や

DOCK といったプログラムを単独又は組み合わせて用いることで、より容易に薬物分子を同定、検索、評価又は設計することができる。また、タンパク質の立体構造座標と薬物分子とのドッキング評価は、例えば前記ステップ II-20 に記載した NEC 社製の BIOCES 等のソフトウェアを用いて行うことができる。

ここで、薬物分子は、既知のものであっても、新たに合成された新規な化学構造を有する薬物分子であっても、その立体構造が得られるものであれば、いずれの薬物分子も本発明の方法で用いることができる。薬物分子の立体構造座標は、X線結晶解析やモデリング等のいずれの方法で得られたものでも良い。3次元構造座標が決定されているものは、適当なデータベース、例えば CCDC (Cambridge Crystallographic Data Centre: <http://www.ccdc.cam.ac.uk/>) や PDB (Protein Data Bank: <http://www.rcsb.org/pdb/>) 等から取得することができる。

更には、標的タンパク質の立体構造を用いて、例えば特開 2000-178209 号公報に記載されている方法によっても、薬物分子を設計することができる。この様に、本発明の方法で得られたタンパク質の立体構造座標を用いることで、薬物分子のコンピュータによる分子設計が可能となる。ただし、本発明の分子設計方法は、これらのプログラムや手法を用いるものに限定されるものではない。

薬物の分子設計には、通常、概念的に 2 つの段階がある。最初の段階は、リード化合物を見つけだすものであり、次の段階はリード化合物の最適化である。どちらの段階も、標的タンパク質の立体構造座標を使用して、それ自体既知の方法により行うことができる。これにより最適な医農薬候補分子を得ることができる。

VI. 分子設計方法により得られる医農薬候補分子のスクリーニング方法

上記方法により同定、検索、評価又は設計された医農薬候補分子は、その分子の性質に応じて、例えばそれ自体既知の化学合成法により得ることができる。しかしながら、薬物分子は、天然化合物、合成化合物のいずれでも良く、また、高分子化合物、低分子化合物のいずれでも良い。得られた医農薬候補分子は、更に、それ自体既知の方法により、試験管内や生体内における薬理学的または

生理学的試験によりその活性を調べ、所望の活性を有する医農薬候補分子を選抜することにより実際に医農薬として応用可能なものを得ることができる。

VII. 医農薬組成物の製造方法

上記スクリーニング方法により選択された医農薬等の薬物分子、例えば医薬分子は、それ自体単独で治療対象となる疾患等の患者に投与することができるが、これらの有効成分の1種又は2種以上を混合して投与することもできる。また、薬理学的に許容される製剤用添加物等を用いて該物質を医薬品組成物として製剤化し、これを投与するのが好ましい。例えば、必要に応じて糖衣を施した錠剤、カプセル剤、顆粒剤、細粒剤、散剤、丸剤、マイクロカプセル剤、リポソーム製剤、トローチ、舌下剤、液剤、エリキシル剤、乳剤、懸濁剤等として経口的に、あるいは無菌の水性液もしくは油性液として製造した注射剤や、座剤、軟膏、貼付剤等として非経口的に使用できる。これらは、例えば、該物質を生理学的に認められる担体、香味剤、賦形剤、ベヒクル、防腐剤、安定剤、結合剤などとともに一般に認められた製剤実施に要求される単位用量形態で混和し、充填又は打錠等の当業界で周知の方法を用いて製造することができる。これらの医薬組成物における有効成分量は指示された範囲の適当な容量が得られるようにするものである。

農薬分子について、実際に農薬として使用する場合には、担体若しくは希釈剤、添加剤および補助剤等と公知の方法で混合して、通常農薬として用いられている製剤形態（組成物）、例えば粉剤、粒剤、水和剤、乳剤、水溶剤、フロアブル剤等に調製して使用される。

実施例

以下、実施例を挙げて本発明を更に具体的に説明するが、下記の実施例は、本発明の具体的な認識を得る一助と見なすべきであり、本発明の範囲を何ら制限するものではない。

実施例1 β2アドレナリンレセプターの立体構造の構築

上記発明の実施形態のI-10～I-40で詳述した方法に従って、次の通りヒト由来

β 2アドレナリンレセプターの誘導適合を含めた立体構造を構築した。第7図にフローチャートを示す。

立体構造モデルの構築は、NEC 社製ワークステーション（機種：Express5800/120Rc-2、CPU：PentiumIII 933MHz x 2、OS：Red Hat Linux 6.2J、メモリ：1024Mbytes）を用いて行った。目的とした β 2アドレナリンレセプターのアミノ酸配列は、PIR；<http://www-nbrf.georgetown.edu/pir/> の ID:QRHUB2 より得た。

この β 2アドレナリンレセプターのアミノ酸配列を目的タンパク質の配列として PSI-BLAST (Position-Specific Iterated BLAST) によるアライメントを行った。その際、モチーフプロファイルは、GCRDb；<http://www.gcrdb.uthscsa.edu/> の全配列 892 個を用いた。 β 2アドレナリンレセプターのアミノ酸配列を、SEQ ID No. 1 に示す。

参照するタンパク質の立体構造として、PDB (<http://www.rcsb.org/pdb/>) の ID:1F88(ロドプシン)のB鎖の構造を用い、このB鎖に対してのアライメントを得た。1F88(ロドプシン)のB鎖の配列を SEQ ID No. 2 に、アライメント結果を第8図に示す。1F88(ロドプシン)の結晶格子中にはA鎖及びB鎖よりなるほぼ同一の立体構造を持つ2量体があり、B鎖を参照構造として用いた。またA鎖とB鎖の座標にはそれぞれ大きな欠損があり完全ではなく、前記ステップ I-40 で詳述したモデリング・プログラム FAMS を用いて 1F88 構造のモデリングを行い、構築された立体構造を β 2アドレナリンレセプターの参照タンパク質立体構造とした。

PDB ファイルおよび FAMS では適当な残基に水素原子が付加されないため、この参照タンパク質立体構造の適当な残基に水素原子を発生させ、基準振動解析法の入力座標となる初期原子座標を得た。

前記ステップ I-10～I-20 のとおり、得られた初期原子座標のデカルト座標系による最適化、SS 結合のポテンシャルパラメータの一部をゼロにしてデカルト座標系で再最適化、2面角座標系による基準振動解析法を行い、固有値・固有ベクトルを得た。

この際、パラメータは AMBER の parm89a Rev A を用いた。非結合相互作用のカットオフ値は内側 9.0 Å、外側 10.0 Å とし、1-4 相互作用のパラメータは非結合

相互作用のそれに 1/2 を乗じたものを使用し、誘電率は距離依存型 ($1/r\text{\AA}$) とした。最適化は、Fletcher-Reeves の共役勾配法を用いた。得られた初期原子座標のデカルト座標系による最適化をしたあと、SS 結合の結合角、2 面角のパラメータをゼロにする以外は同じ条件を使用してデカルト座標系で再最適化し、2 面角座標系による基準振動解析法を行い、固有値・固有ベクトルを得た。

使用した最適化の条件は、Sumikawa, H., Suzuki, E.-I., Fukuhara, K.-I., Nakajima, Y., Kamiya, K., and Umeyama H. 1998. Dynamics structure of granulocyte colony-stimulating factor proteins studied by normal mode analysis: Two domain-type motions in low frequency modes. Chem Pharm Bull 46: 1069-1077 に記載されている方法を用いた。また、2 面角座標系による基準振動解析法の詳細は、Noguti, T., and Go, N. 1983. Dynamics of native globular proteins in terms of dihedral angles. J Phys Soc Jpn 52: 3283-3288 および Noguti, T., and Go, N. 1983. A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. J Phys Soc Jpn 52: 3685-3690 に記載されている方法を用いた。

前記ステップ I-30 のとおり、温度を 300° K とし、 30cm^{-1} 以下の各固有値に対する $\text{C}\alpha$ 原子のゆらぎを求め、PDB ID:1F88 (ロドプシン) の A 鎖と B 鎖の平均の温度因子から換算される $\text{C}\alpha$ 原子のゆらぎとの比をとり、各固有値に対する平均の比を得た。平均の比をこの固有値に属する固有ベクトルに掛けて、参照タンパク質の原子座標に加えて変位を行い、誘導適合 (induced fit) 型参照タンパク質の立体構造を規定する座標を得た。同様に固有ベクトルに -1 を掛けた変位、2 倍した平均の比を固有ベクトルに掛けた変位、さらに -1 を掛けた変位を行った。ただし、ここで加える固有ベクトルは 2 面角座標からデカルト座標に変換してある。1 つの固有値・固有ベクトルから 4 つの誘導適合 (induced fit) 型参照タンパク質の立体構造セットが得られる。用いた 30cm^{-1} 以下の固有値の数は 118 個であり、得られた誘導適合 (induced fit) 型の参照タンパク質の数は 472 個である。例として、第 9 図に最低固有値 4.47cm^{-1} の M^* (=26.4) 倍したゆらぎと換算した温度因子を示す。

前記ステップ I-40 のとおり、非誘導適合 (no induced fit) 型参照タンパク質

立体構造と誘導適合 (induced fit) 型参照タンパク質立体構造セットから FAMS により目的タンパク質である β 2 アドレナリンレセプターの立体構造をモデリングした。目的タンパク質の立体構造と参照タンパク質の立体構造は 1 対 1 の関係にあり、472 個の誘導適合 (induced fit) 型目的タンパク質立体構造と従来の方法から得られる 1 個の非誘導適合 (no induced fit) 型目的タンパク質立体構造を得た。例として、第 10 図に、上記で得られた非誘導適合 (no induced fit) 型参照タンパク質から構築された非誘導適合 (no induced fit) 型目的タンパク質立体構造と最低固有値の固有ベクトルを $\pm 2 \times M^{\circ}$ ($\pm 2 \times 26.4$) 倍した誘導適合 (induced fit) 型参照タンパク質立体構造から構築された誘導適合 (induced fit) 型目的タンパク質の立体構造の一部を示す。図中、中央の構造が非誘導適合型目的タンパク質である。

実施例 2 トリプシン単体およびトリプシン・インヒビター単体からの複合体の立体構造の構築

本例では受容体、リガンド、受容体-リガンド複合体の X 線結晶解析が既知である牛肺臓由来の β -Trypsin(トリプシン)とトリプシン・インヒビター(BPTI)の系を用いて、本発明のタンパク質-リガンド複合体の立体構造構築方法の検証を行った。ここではトリプシンが受容体タンパク質(目的タンパク質)、BPTI がリガンドである。

用いたトリプシンのアミノ酸配列を SEQ ID No. 3 に、トリプシン・インヒビター(BPTI)のアミノ酸配列を SEQ ID No. 4 に示す。なお、トリプシンのアミノ酸番号は、キモトリプシノーゲン(キモトリプシンの前駆体)のアミノ酸配列番号で記述するので、次に示す通り、アミノ酸番号 16~245 までの 223 残基になる。途中、アミノ酸番号 35、36、68、128、131、188、205、206、207、208 に欠落が、184、188、221 に重複(184A、188A、221A で表示)がある。

Ile	Val	Gly	Gly	Tyr	Thr	Cys	Gly	Ala	Asn	Thr	Val	Pro	Tyr	Gln	Val
16															30
Ser	Leu	Asn	Ser	Gly	Tyr	His	Phe	Cys	Gly	Gly	Ser	Leu	Ile	Asn	Ser
34	37														45

Gln Trp Val Val Ser Ala Ala His Cys Tyr Lys Ser Gly Ile Gln Val
 50 55 60 65
 Arg Leu Gly Glu Asp Asn Ile Asn Val Val Glu Gly Asn Glu Gln Phe
 67 69 70 75 80
 Ile Ser Ala Ser Lys Ser Ile Val His Pro Ser Tyr Asn Ser Asn Thr
 85 90 95
 Leu Asn Asn Asp Ile Met Leu Ile Lys Leu Lys Ser Ala Ala Ser Leu
 100 105 110
 Asn Ser Arg Val Ala Ser Ile Ser Leu Pro Thr Ser Cys Ala Ser Ala
 115 120 125 127 130 132
 Gly Thr Gln Cys Leu Ile Ser Gly Trp Gly Asn Thr Lys Ser Ser Gly
 135 140 145
 Thr Ser Tyr Pro Asp Val Leu Lys Cys Leu Lys Ala Pro Ile Leu Ser
 150 155 160
 Asp Ser Ser Cys Lys Ser Ala Tyr Pro Gly Gln Ile Thr Ser Asn Met
 165 170 175 180
 Phe Cys Ala Gly Tyr Leu Glu Gly Lys Asp Ser Cys Gln Gly Asp
 183 184A184 185 187 188A188 190
 Ser Gly Gly Pro Val Val Cys Ser Gly Lys Leu Gln Gly Ile Val Ser
 195 200 204 209 210
 Trp Gly Ser Gly Cys Ala Gln Lys Asn Lys Pro Gly Val Tyr Thr Lys
 215 217 219 220 221A221 225 230
 Val Cys Asn Tyr Val Ser Trp Ile Lys Gln Thr Ile Ala Sse Asn
 235 240 245

前記ステップ II-10～II-30で詳述した方法に従って、次の手順でトリプシン-BPTI 褐合体の立体構造モデルを構築し、褐合体活性部位の位置をその X 線結晶解析データと比較検討した。

受容体タンパク質-リガンド褐合体の立体構造モデルの構築は、DEL 社製パーソナルコンピュータ(機種: Dimension XPS B866、CPU: Pentium III 864MHz、OS:

RedHat Linux 6.2J、メモリ：512Mbytes)を用いて行った。トリプシンと BPTI 単独のX線結晶解析の座標、ならびにトリプシン-BPTI 複合体のそれは、Protein Data Bank (PDB)；<http://www.rcsb.org/pdb/>より、それぞれ 1TLD(トリプシン単体)、4PTI(BPTI)、2PTC(トリプシン-BPTI 複合体)を取得して用いた。

トリプシンと BPTI の立体座標系は、トリプシン-BPTI 複合体の結果を考察しやすいように 1TLD と 4PTI の座標系を 2PTC の座標系に最小二乗法によりスーパーインポーズした。トリプシンと BPTI の立体座標はヘテロ原子に水素原子を発生させたのち、それぞれ単体での初期座標の最適化を行った。次にトリプシンは BPTI を含まない系で基準振動解析を行い、波長ごとにに対する振動ベクトルを求めた。

その中で、時間的に長周期な振動ベクトルからなる 5 つのトリプシンの立体構造に対して、BPTI の立体構造をドッキングして apricot-MCSS プログラムによる MCSS 計算を行い、トリプシン-BPTI 複合体の立体構造を精密化した。MCSS 計算の内訳は、最初に 1000 ステップのトリプシン-BPTI 複合体の分子力学計算による立体構造の最適化を行い、続いて 1fs を 1 ステップとする 300° K、10ps の分子動力学計算によりトリプシン-BPTI 複合体の立体構造の緩和を行った。分子動力学計算では複合体の立体構造が大きく崩れないように式(27)に示した C_α 原子に対する K_{xyz} = 10.0 kcal/mol/Å² の拘束条件を加えた。そして 10ps 後の立体構造について、トリプシン-BPTI 複合体の座標データを PDB フォーマットで得た。

MCSS 計算後のトリプシン-BPTI 複合体系のトリプシンの立体構造を第 1-1 図に示す。トリプシンの原子座標を眺めて見ると、主鎖、側鎖ともに大きくばらついている部分と、それらが余りばらついていない部分があった。その中でもトリプシン活性部位であるトリプシン側の His57、Asp102、Gly193-Asp194-Ser195 (オキシアニオンホール) 部分は主鎖、側鎖ともよく一致していた。このことを利用するとリガンド結合部位に重要な受容体タンパク質側の部位を見つけられる。それは新たなリガンドをデザインする上でたいへん参考になる。

MCSS 計算前のトリプシン-BPTI 複合体の初期立体構造を第 1-2 図に、MCSS 計算後のトリプシン-BPTI 複合体の立体構造を第 1-3 図に、複合体の X 線結晶解析の立体構造とともに示した。これらの図では、トリプシン-BPTI 複合体の活性部

位に当たる、トリプシン側では His57、Asp102、オキシアニオンホール(Gly193-Asp194-Ser195)を、BPTI 側では Lys15だけを抜き出して表示した。黒色で表示されている線がトリプシン-BPTI 複合体の X 線結晶解析の立体構造、灰色で表示されている線が本発明により組み立てられた複合体モデルの初期の立体構造(第12図)と精密化された結果(第13図)である。

トリプシンの活性部位である His57、Asp102、オキシアニオンホールは、MCSS 計算前の初期立体構造(第12図)と MCSS 計算後の精密化された立体構造(第13図)は主鎖、側鎖を含めてよく一致している。BPTI の Lys15 主鎖も、そのカルボニル酸素がオキシアニオンホールの Gly193 と Ser195 ベプチド NH 基と 2 本の水素結合で結ばれているため、MCSS の計算前後でよく一致している。一方 BPTI の Lys15 側鎖の方向は、MCSS 計算前はトリプシンの活性ポケットに入っていないが、MCSS 計算で立体構造を精密化することによりその活性ポケットに入り込み、トリプシン-BPTI 複合体の X 線結晶解析によく一致するようになる。

このことは、目的タンパク質の基準振動モードを含む複数のモデル立体構造を用いること、それらにドッキングして得られる目的タンパク質-リガンド複合体の初期立体構造を MCSS 計算によりシミュレーションする手法が、目的とするタンパク質-リガンド複合体の立体構造の構築に有用であることを示している。

実施例 3 トリプシン、トリプシン・インヒビターそれぞれの結合部位の特定

前記ステップ III-10～III-30 で詳述した方法に従って、次の手順でトリプシンおよび BPTI の結合部位をそれぞれ特定し、それら部位を複合体の X 線結晶解析データと比較検討した。本例では、タンパク質-リガンド複合体 X 線結晶解析が既知である牛臍臓由来の β -Trypsin(トリプシン)とトリプシン・インヒビター(BPTI)の系を用いた。ここではトリプシンが受容体タンパク質(目的タンパク質)、BPTI がリガンドであるが、BPTI もタンパク質であるので、タンパク質側だけでなく、リガンド側の結合部位の特定も行った。用いたトリプシンおよびトリプシン・インヒビター(BPTI)アミノ酸配列は、それぞれ SEQ ID No. 3 および SEQ ID No. 4 に示した通りである。

トリプシン-BPTI 複合体の立体構造座標は、Protein Data Bank (PDB);

<http://www.rcsb.org/pdb/>より 2PTC を得た。2PTC のトリプシン-BPTI 複合体の X 線結晶解析の立体構造を第 14 図に示す。

タンパク質ならびにリガンドの結合部位の検索には、DEL 社製パーソナルコンピュータ(機種: Dimension XPS B866、CPU: Pentium III 864MHz、OS: RedHat Linux 6.2J、メモリ: 512Mbytes)を用いた。

トリプシンと BPTI の立体構造座標はそれぞれ別に扱い、ヘテロ原子に水素原子を発生させたのち、周囲に水溶媒を発生した。次にトリプシンと BPTI の中で MSAS が 30%以上のアミノ酸残基が形成する表面より 3.5 Å 以内の水分子をベンゼン分子と置換した。その際ベンゼン同士が 1.5 Å 以内になるときは水分子のベンゼンへの置換は行わなかった。そしてベンゼンへの置換が終了した時点で水分子は 1 回消去した。ベンゼン分子を含むトリプシンと BPTI の立体構造座標はそれら周囲に水分子を満たした周期ボックスを発生させたのち、水分子の周期境界条件のもと apricot プログラムによる経験的分子エネルギー計算を実行した。これらエネルギー計算の内訳は最初に 1,000 ステップの分子力学計算による構造の最適化、続いて 1fs を 1 ステップとする 300° K、10ps の分子動力学計算によるベンゼン分子の挙動探索である。分子動力学計算ではタンパク質の立体構造が大きく崩れないよう全アミノ酸残基の C_α 原子に式(27)による $U_{xyz}=10.0 \text{ kcal/mol}/\text{\AA}^2$ の拘束条件を加えた。

これら経験的分子エネルギー計算の終了した時点で、トリプシン、BPTI とともに周期ボックス内の水分子を消去し、分子動力学計算 10ps 後のトリプシンとベンゼンの原子座標および BPTI とベンゼンの原子座標を PDB フォーマットで得た。それからトリプシンならびに BPTI を除いたベンゼンの分布について閾値を 6 Å としたクラスター解析をそれぞれ行った。トリプシンと BPTI 周囲にそれぞれ置かれた 94 個と 40 個のベンゼン分子のうち、一番大きなクラスターはそれぞれ 29 個、11 個であった。トリプシンと BPTI 周囲のベンゼン分子の分布を、トリプシンと BPTI とともに第 15 図と第 16 図にそれぞれ示す。

これらの図は、第 14 図と同じ方向から見たものである。図中、黒線の六角形が一番大きなベンゼンクラスターである。

第 14 図～第 16 図より、トリプシンと BPTI 周囲の一番大きなベンゼンクラス

ター同士は方向的によく一致していることが分かる。すなわちタンパク質の疎水性残基の周囲にベンゼン分子を配置し、水溶媒中での分子動力学計算を行い、クラスター解析による大きなベンゼンクラスター分布を探索することにより、タンパク質のリガンドへの結合部位候補を特定できることが分かる。またグラフィックス上でこれらのクラスター同士を重ねるようにタンパク質とリガンドをドッキングさせると、タンパク質ーリガンド複合体の初期立体配置をラフに予測できると考えられる。この初期立体配置は手動あるいは分子設計ソフトで調整することにより、タンパク質ーリガンド複合体の立体配置の有力な候補の1つになる。

産業上の利用可能性

上記のとおり、本発明の方法は、従来の方法と比べて、より真に近いタンパク質の構造、特にリガンドと結合する近傍を精度良く構築しうる方法である。したがって、本発明の方法は医農薬分子の設計等に極めて有用である。

即ち、本発明の誘導適合を含めた立体構造の構築方法は、目的タンパク質のモデル立体構造による基準振動解析から得られる複数の座標データを用いるものであり、分子振動を考慮した平均のモデル立体構造が精度よく構築できる。とくに目的タンパク質ーリガンド複合体の立体構造を予測する場合には、それに重要な誘導適合 (induced fit) を含められるので、それを考慮した精密な複合体のモデル立体構造を構築できる。また複数の受容体タンパク質の立体構造を1つのリガンドのそれで構造最適化させる Multiple Copy Simultaneous Search (MCSS) 法でタンパク質ーリガンド複合体の立体構造をシミュレートすることにより、経時的に平均化された複合体の立体構造が得られる。

また、本発明のタンパク質ーリガンド複合体の立体構造構築方法は、MCSS 計算後に、目的タンパク質ーリガンド複合体モデルにおける受容体側の原子座標のばらつきを調べるものであり、活性に重要なサイトは原子座標のばらつきが比較的小さく、その他のサイトはそのばらつきが大きいことをを利用して、新たにリガンドをデザインすることができ、医農薬分子設計において、有効に利用することができます。

本発明を詳細にまた特定の実施態様を参照して説明したが、本発明の精神と範囲を逸脱することなく様々な変更や修正を加えることができることは当業者にとって明らかである。

本出願は、2001年1月19日の日本特許出願（特願2001-011783号）に基づくものであり、その内容はここに参考として取り込まれる。また、本明細書にて引用した文献の内容もここに参考として取り込まれる。

請求の範囲

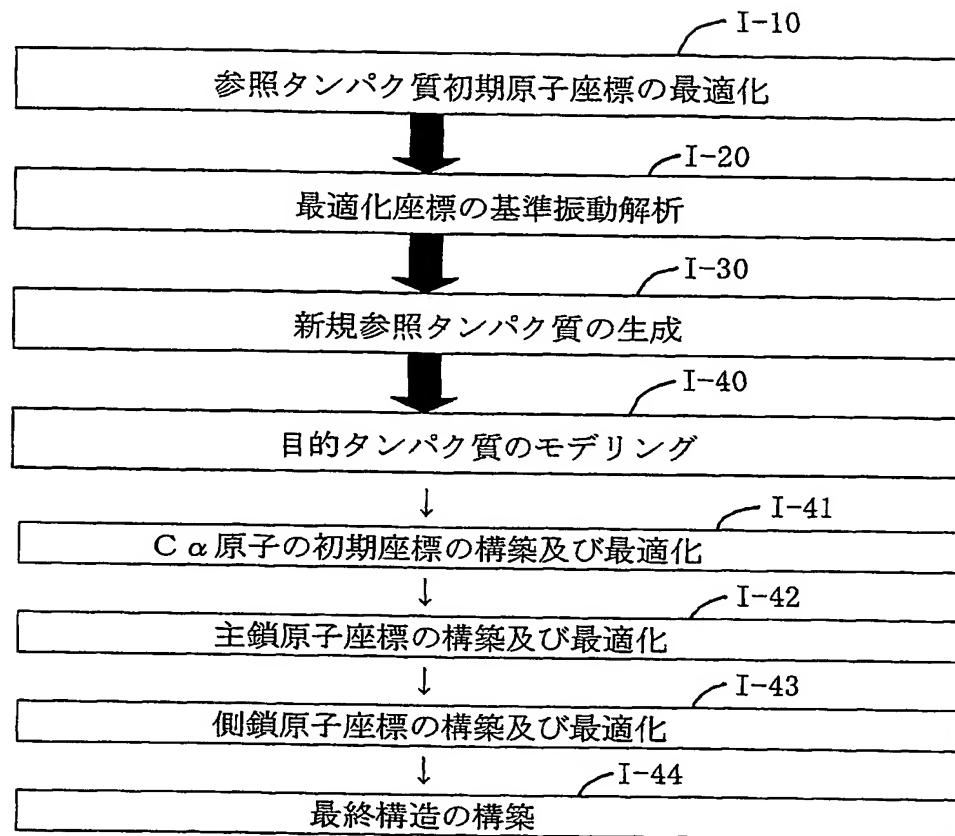
1. 参照タンパク質と目的タンパク質とのアライメントを導き出し、該アライメントおよび参照タンパク質の立体構造情報に基づいて目的タンパク質の立体構造を構築する方法において、参照タンパク質の立体構造とその原子座標を変位させた複数の立体構造を参照タンパク質の立体構造として目的タンパク質の複数の立体構造セットを作成することを特徴とする誘導適合を含めたタンパク質の立体構造構築方法。
2. 参照タンパク質の原子座標の変位が、基準振動解析法により行われることを特徴とする請求項 1 に記載の方法。
3. 立体構造の構築が、(i) アミノ酸中の $C\alpha$ 原子について参照タンパク質の立体構造から座標を取得し、目的関数を最小化するように $C\alpha$ 原子座標を最適化し、(ii) 最適化された $C\alpha$ の原子座標に主鎖の他の原子を付加して目的関数を最小化するように主鎖の原子座標を最適化し、(iii) 最適化された主鎖の原子座標に側鎖の他の原子を付加し目的関数を最小化するように最適化することにより行われることを特徴とする請求項 1 又は 2 に記載の方法。
4. (i) 請求項 1 ~ 3 のいずれかに記載の方法により得られる目的タンパク質の複数の立体構造とリガンドとのドッキング操作を行い、
(ii) 目的タンパク質の 1 つの構造とリガンドとの構造の経験的分子エネルギー計算を、目的タンパク質の構造の数だけ行い、その際、
(iii) 目的タンパク質側は、複数の構造それぞれのポテンシャルエネルギー勾配に応じて原子座標を動かし、
(iv) リガンド側は、複数個算出されたポテンシャルエネルギー勾配を平均化した方向にリガンドの原子座標を動かして、
(v) 目的タンパク質の複数の立体構造に基づくリガンドの立体構造を求めることを特徴とするタンパク質-リガンド複合体の立体構造構築方法。
5. 経験的分子エネルギー計算において、目的タンパク質の初期 $C\alpha$ 原子座標の位置をオプション Harmonic 関数として加えるか、あるいは目的タンパク質の主鎖のねじれ角を拘束するポテンシャル関数を加えることを特徴とする請求項 4 に記

載の方法。

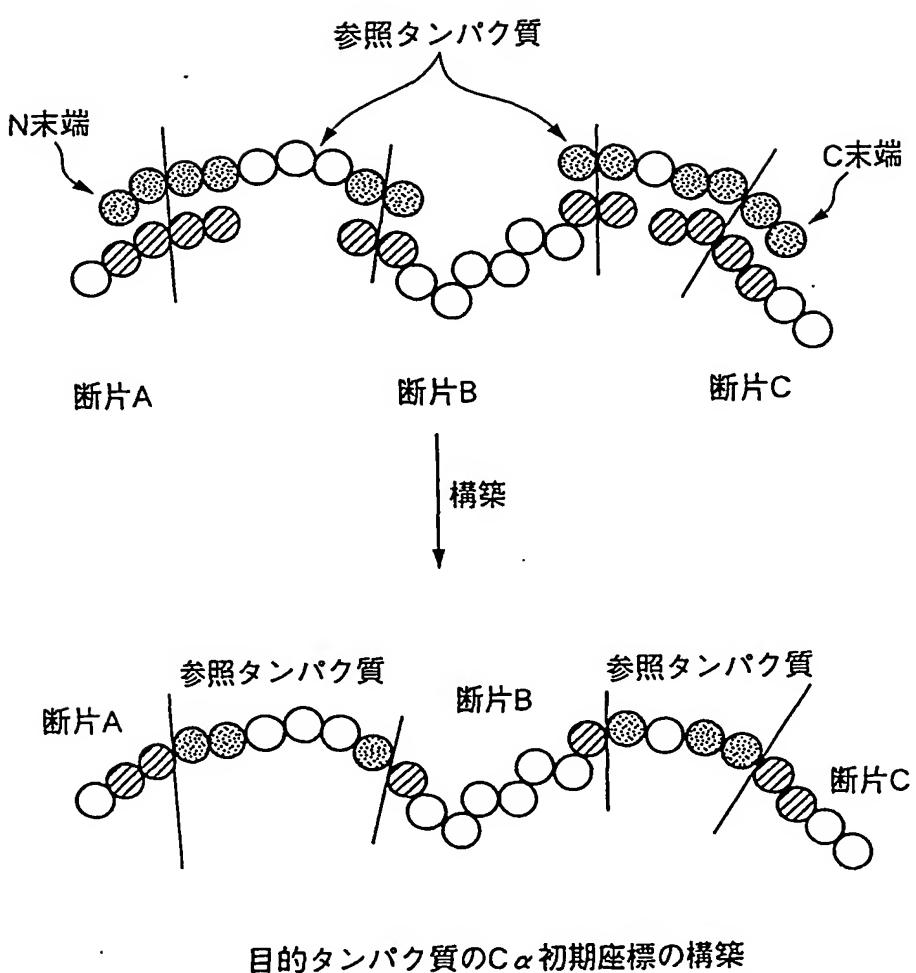
6. (i) タンパク質の立体構造の周囲に低分子化合物を配置し、
(ii) それらの周囲にさらに水分子を配置し、水溶媒中での経験的分子エネルギー計算を行って、タンパク質と低分子化合物との原子座標を得、
(iii) 得られた原子座標について、タンパク質の周囲および内部の、低分子化合物の挙動解析を行い、リガンドの結合部位を判定することを特徴とするタンパク質のリガンド結合部位の特定方法。
7. (i) タンパク質およびリガンドの立体構造の周囲に低分子化合物を配置し、
(ii) それらの周囲にさらに水分子を配置し、水溶媒中での経験的分子エネルギー計算を行って、タンパク質と低分子化合物との原子座標を得、
(iii) 得られた原子座標について、タンパク質およびリガンドの周囲および内部の、低分子化合物の挙動解析を行い、タンパク質—リガンド複合体の結合部位を判定することを特徴とするタンパク質—リガンド複合体の結合部位の特定方法。
8. 低分子化合物の挙動解析が、低分子化合物を対象としたクラスター解析により行われ、得られたクラスターのサイズをリガンドの結合可能性部位の順位として結合部位を判定することを特徴とする請求項 6 または 7 に記載の方法。
9. 請求項 6 ~ 8 のいずれか 1 項に記載の方法により特定したタンパク質のリガンド結合部位にリガンドをドッキングし、経験的分子エネルギー計算によりタンパク質—リガンドの複合体の立体構造を得ることを特徴とするタンパク質—リガンド複合体の立体構造構築方法。
10. 請求項 1 ~ 5 および 9 のいずれか 1 項に記載の方法により得られるタンパク質の立体構造および／またはタンパク質—リガンド複合体の立体構造を規定する原子座標が記録されていることを特徴とするコンピューター読みとり可能な記録媒体。
11. 請求項 1 ~ 5 および 9 のいずれか 1 項に記載の方法により得られるタンパク質の立体構造および／またはタンパク質—リガンド複合体の立体構造を規定する原子座標を含むことを特徴とするデータベース。
12. 請求項 10 に記載の記録媒体または請求項 11 に記載のデータベースか

ら得られるタンパク質の立体構造を規定する原子座標を用いて、薬物候補分子の立体構造との相互作用に基づいて、目的とする薬物分子を同定、検索、評価または設計することを特徴とする薬物分子設計方法。

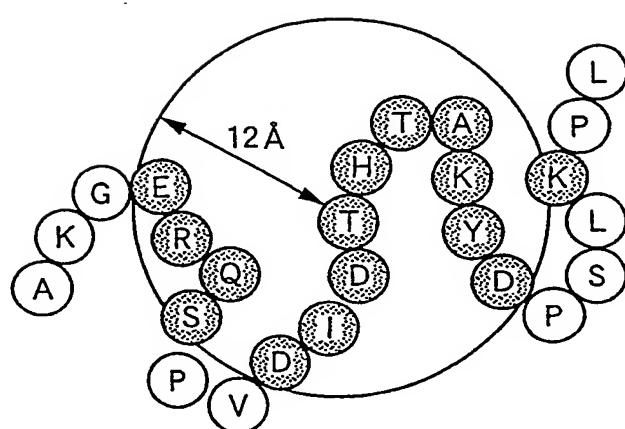
第1図



第2図

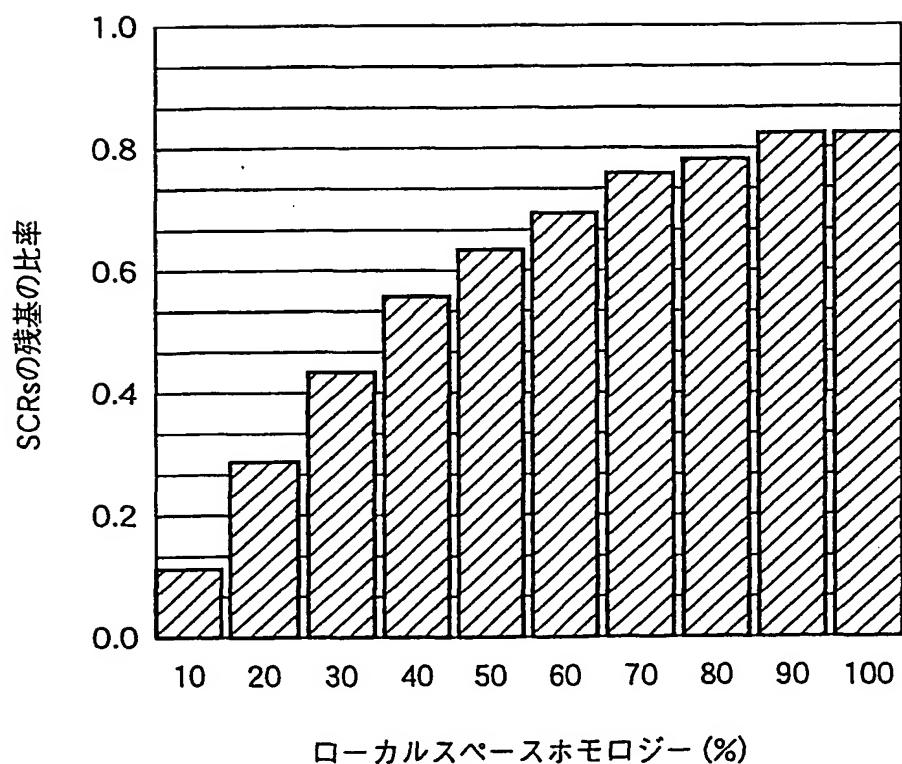


第3図

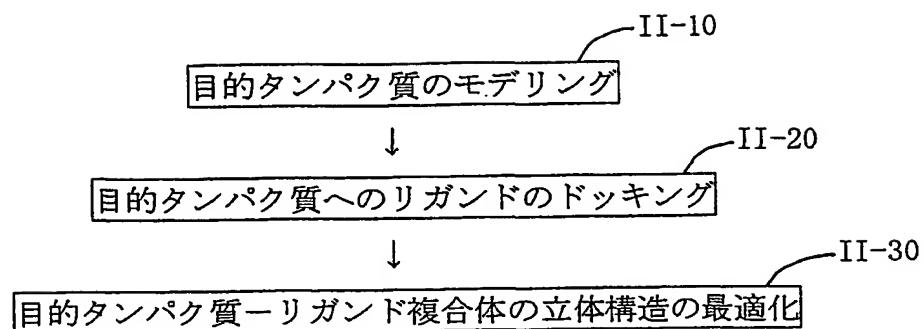


AKG	ERQ-S	PV	DIDTHTAKYD	PSL	K	PL
ARG	NREAS	SP	NVDTHSARYD	--L	K	PL
*	*	*	*	***	*	***

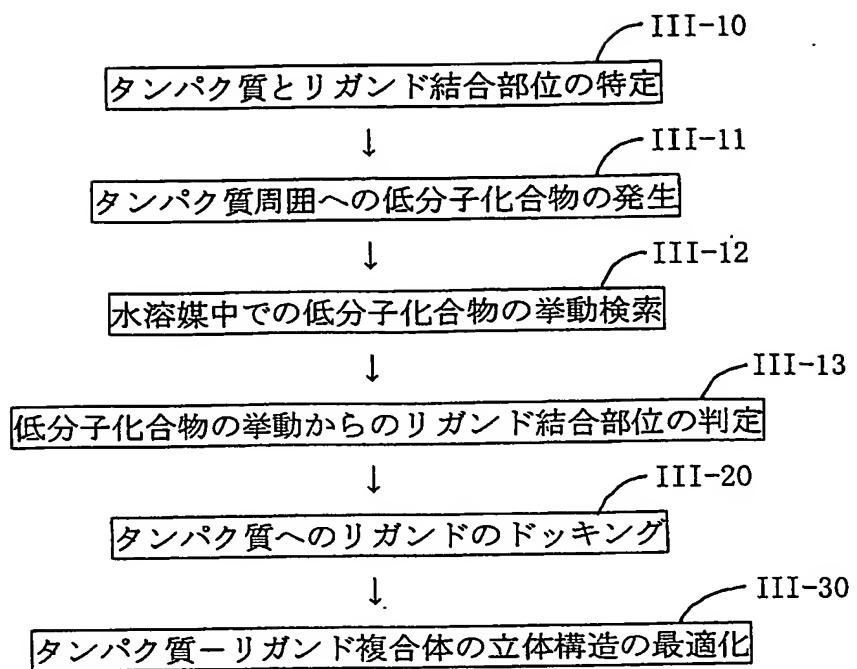
第4図



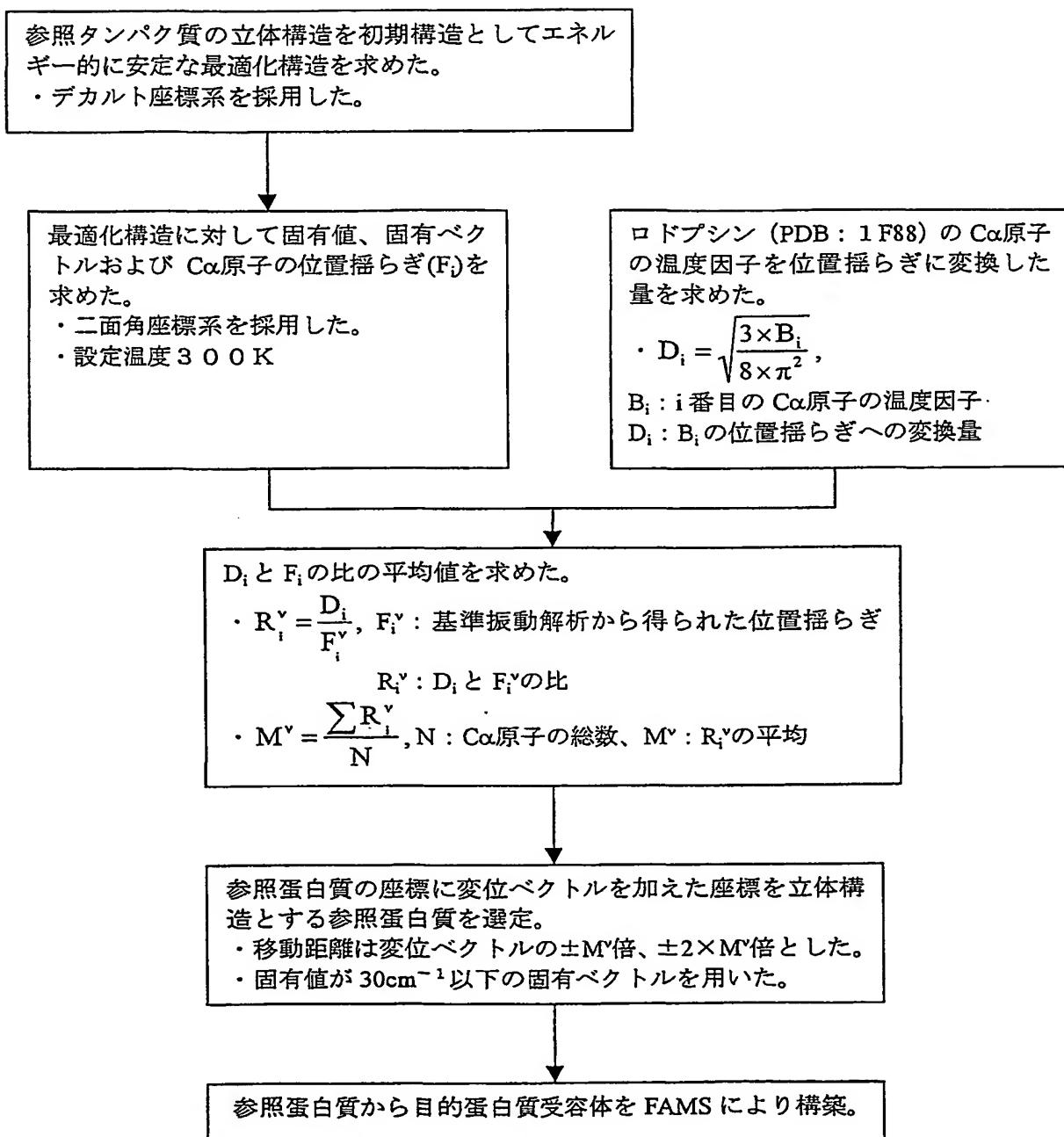
第5図



第6図



第7図



第8図

	354	HOMOLOGY	MATCH	MISSMATCH	INSERTION	DELETION
>QRHUB2	332	19.6%	65	262	5	27

FLLAPNRSHADPHDVTQQRDEVVVV-GMGIVMSLIVLAIIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVMGLAVVP
 FSNKTKGVVRSPFEAPQQYLAEPWQFSMLAAYMFLIIMLGFPINFLTLVTVQHKKLRTPLNYILLNLAVALDFMFVFGFT
 +++++++++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*

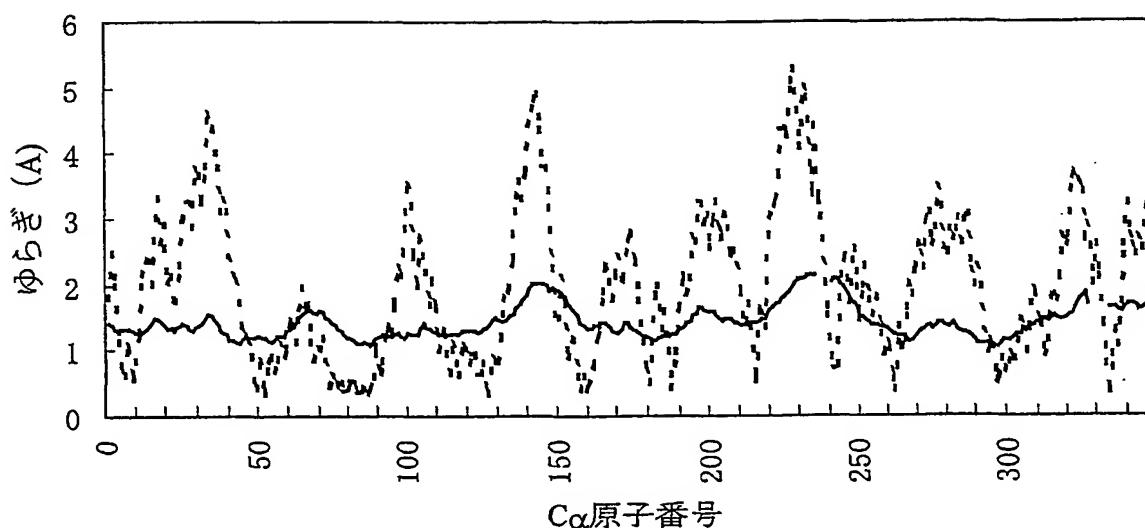
 FGAAHILMKWTFGNFWCEFWTSIDVLCVTASIEETLCVIAVDRYFAITSPFKYQSLLTKNKA
 TTLYTSLHGIFYFGPTGCNLEGFFATLGGEIALWSLVVLAIREYVVVCKPM-SNFRFGENHAIMGVAFTTWVMA
 LACAPP
 +++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*

 IQMHWYRATH-QEAINCYANE-TCCDFFTNQAYAIASSIVSFYYVPLVIMVFSRVFQEAKRQLQKIDKSEGFRHVQNL
 LV-GWSRYIPEGMQCSCGIDYYPHEETNNESFVIMFVVFHFIPLIVIFFCYGQLVFTVKEAAAQQQESAT-----
 +++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*

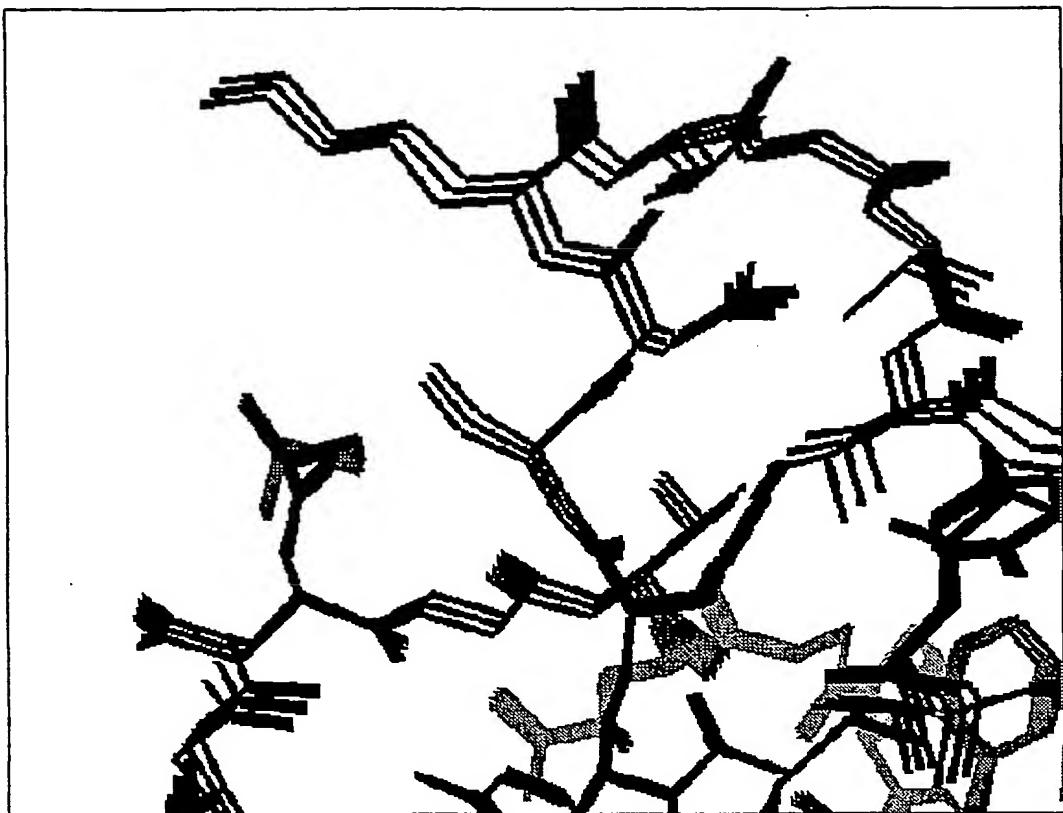
 QVEQDGRTGHGLRRSSKFCLEHKALKTLGIIMGTFTLCWLPPFFIVNIVHHIQDNLI-RKEVYILLNWIGYVNNSGFNPLI
 -----TQKAKEVTRMVIMVIAFLICWLPHYAGVAFYIFTHQGSDFGPIMFTIPAFFAKTSAVYNPV
 ++++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*+++++*

 Y-CRSPDFRIAFQELLCLRRSSLKAYGNGYSSNGNTGEQ
 YIMMNKQFRNCMVTTLCCGKNPPLGDDAEASTTVSKTETSQ
 ++++++++++*+++++*+++++*+++++*+++++*

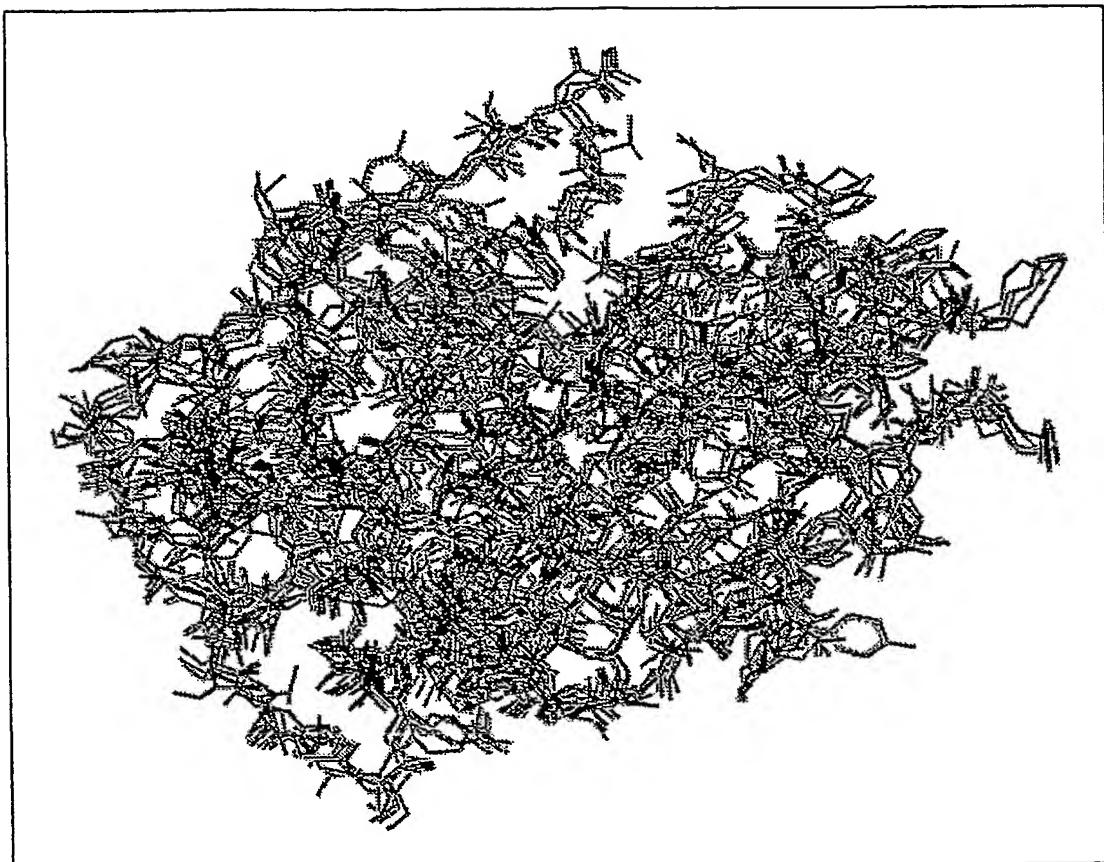
第9図



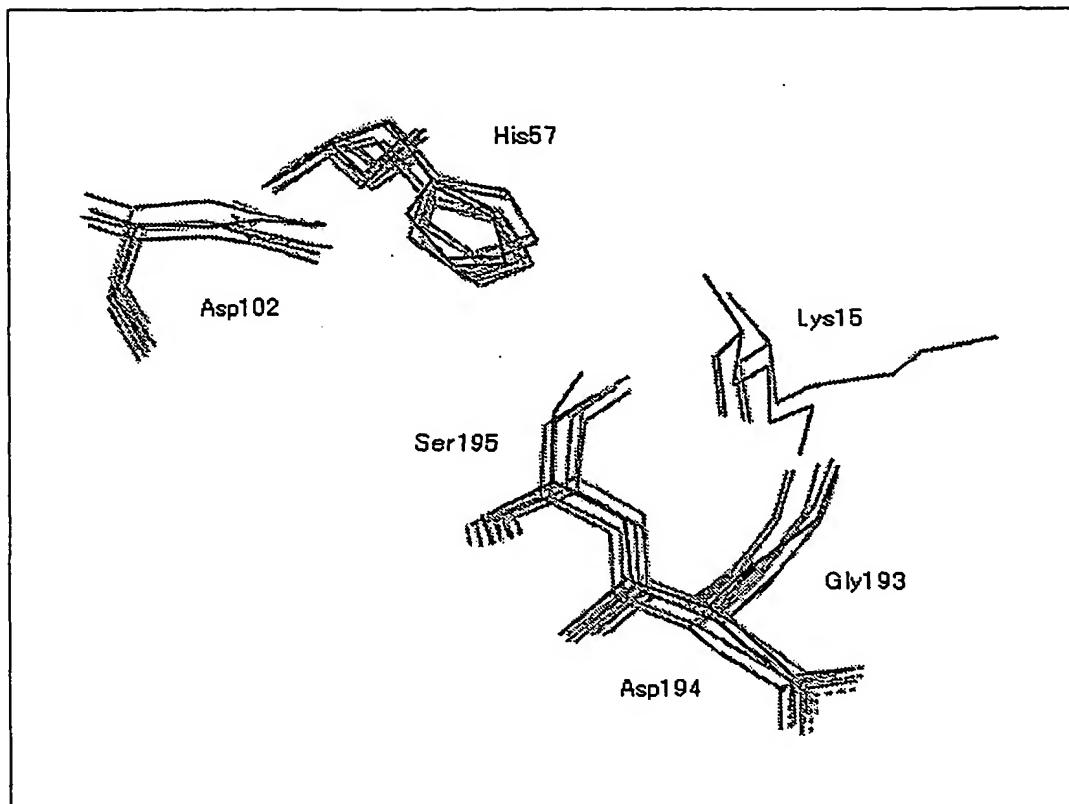
第10図



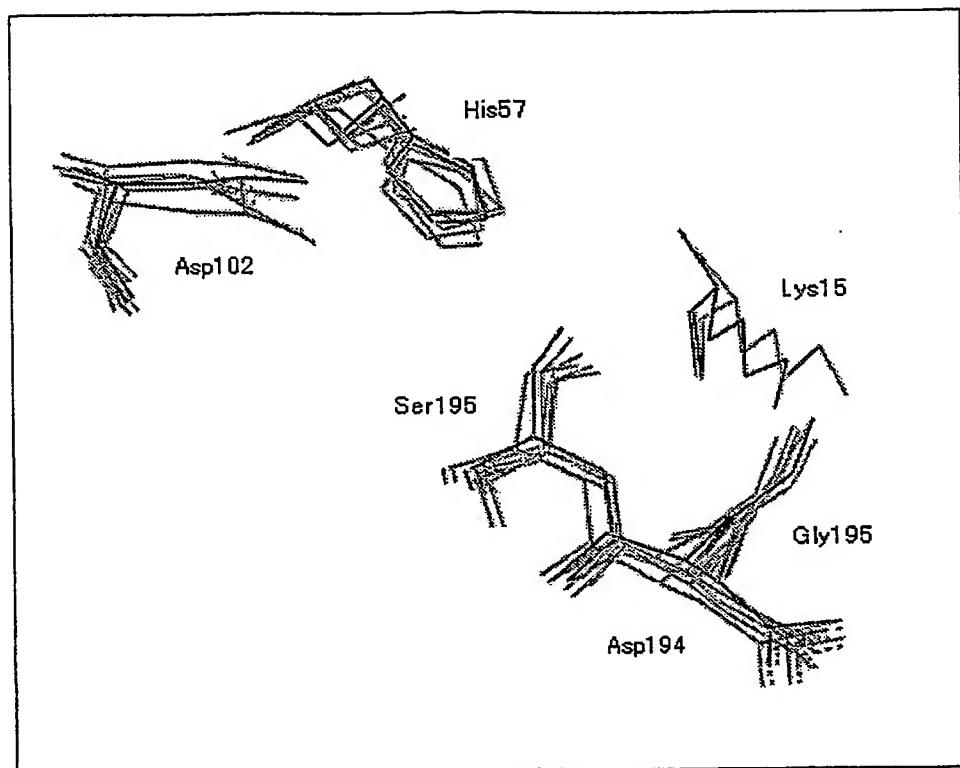
第 11 図



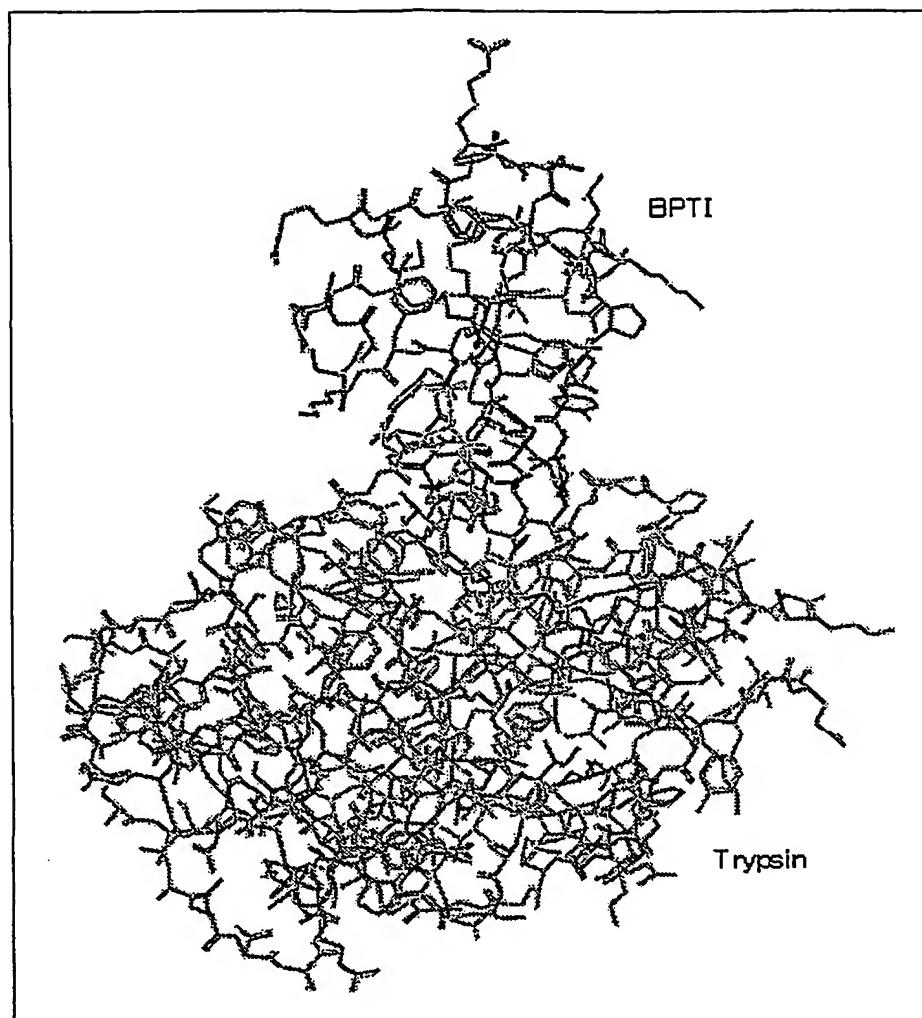
第12図



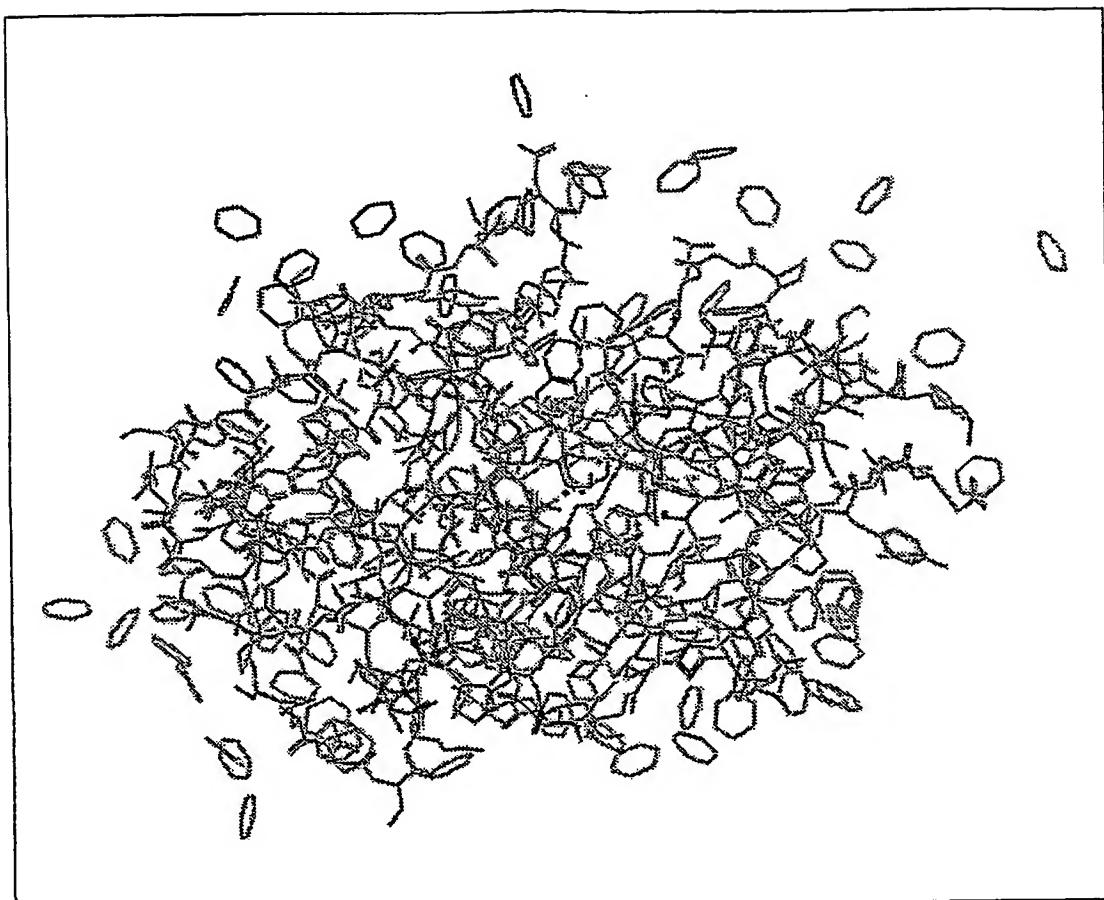
第13図



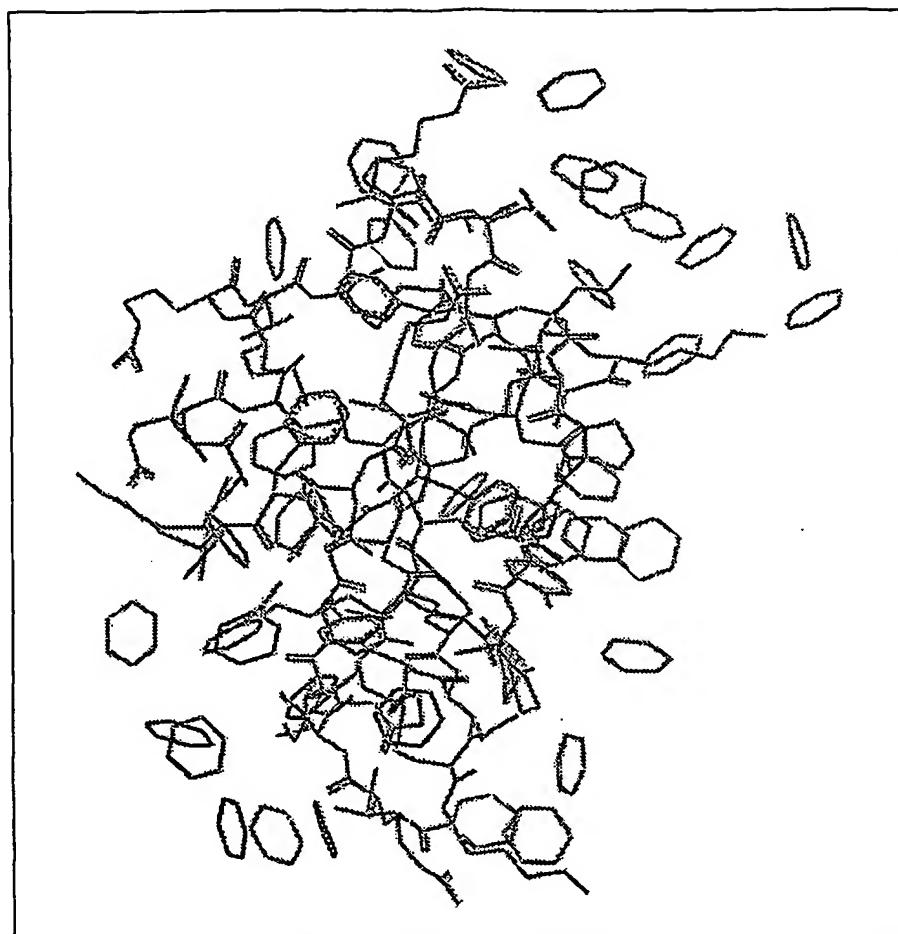
第14図



第15図



第 16 図



SEQUENCE LISTING

<110> Umeyama, Hideaki

<110> Mitsubishi Chemical Corporation

<120> A method for determination of a three-dimensional structure containing induced fit and the use thereof

<130> P-39949

<150> JP 2001-011783

<151> 2001-01-19

<160> 4

<210> 1

<211> 413

<212> PRT

<213> Homo sapiens

<220>

<223>

<400> 1

Met Gly Gln Pro Gly Asn Gly Ser Ala Phe Leu Leu Ala Pro Asn Arg

1 5 10 15

Ser His Ala Pro Asp His Asp Val Thr Gln Gln Arg Asp Glu Val Trp

20 25 30

Val Val Gly Met Gly Ile Val Met Ser Leu Ile Val Leu Ala Ile Val

35 40 45

Phe Gly Asn Val Leu Val Ile Thr Ala Ile Ala Lys Phe Glu Arg Leu

50 55 60

Gln Thr Val Thr Asn Tyr Phe Ile Thr Ser Leu Ala Cys Ala Asp Leu

65 70 75 80

Val Met Gly Leu Ala Val Val Pro Phe Gly Ala Ala His Ile Leu Met

85 90 95

Lys Met Trp Thr Phe Gly Asn Phe Trp Cys Glu Phe Trp Thr Ser Ile
100 105 110
Asp Val Leu Cys Val Thr Ala Ser Ile Glu Thr Leu Cys Val Ile Ala
115 120 125
Val Asp Arg Tyr Phe Ala Ile Thr Ser Pro Phe Lys Tyr Gln Ser Leu
130 135 140
Leu Thr Lys Asn Lys Ala Arg Val Ile Ile Leu Met Val Trp Ile Val
145 150 155 160
Ser Gly Leu Thr Ser Phe Leu Pro Ile Gln Met His Trp Tyr Arg Ala
165 170 175
Thr His Gln Glu Ala Ile Asn Cys Tyr Ala Asn Glu Thr Cys Cys Asp
180 185 190
Phe Phe Thr Asn Gln Ala Tyr Ala Ile Ala Ser Ser Ile Val Ser Phe
195 200 205
Tyr Val Pro Leu Val Ile Met Val Phe Val Tyr Ser Arg Val Phe Gln
210 215 220
Glu Ala Lys Arg Gln Leu Gln Lys Ile Asp Lys Ser Glu Gly Arg Phe
225 230 235 240
His Val Gln Asn Leu Ser Gln Val Glu Gln Asp Gly Arg Thr Gly His
245 250 255
Gly Leu Arg Arg Ser Ser Lys Phe Cys Leu Lys Glu His Lys Ala Leu
260 265 270
Lys Thr Leu Gly Ile Ile Met Gly Thr Phe Thr Leu Cys Trp Leu Pro
275 280 285
Phe Phe Ile Val Asn Ile Val His Val Ile Gln Asp Asn Leu Ile Arg
290 295 300
Lys Glu Val Tyr Ile Leu Leu Asn Trp Ile Gly Tyr Val Asn Ser Gly
305 310 315 320
Phe Asn Pro Leu Ile Tyr Cys Arg Ser Pro Asp Phe Arg Ile Ala Phe

	325	330	335
Gln Glu Leu Leu Cys Leu Arg Arg Ser Ser Leu Lys Ala Tyr Gly Asn			
	340	345	350
Gly Tyr Ser Ser Asn Gly Asn Thr Gly Glu Gln Ser Gly Tyr His Val			
	355	360	365
Glu Gln Glu Lys Glu Asn Lys Leu Leu Cys Glu Asp Leu Pro Gly Thr			
	370	375	380
Glu Asp Phe Val Gly His Gln Gly Thr Val Pro Ser Asp Asn Ile Asp			
	385	390	395
Ser Gln Gly Arg Asn Cys Ser Thr Asn Asp Ser Leu Leu			400
	405	410	

<210> 2

<211> 348

<212> PRT

<213> Bos taurus

<220>

<223>

<400> 2

Met Asn Gly Thr Glu Gly Pro Asn Phe Tyr Val Pro Phe Ser Asn Lys

1	5	10	15
---	---	----	----

Thr Gly Val Val Arg Ser Pro Phe Glu Ala Pro Gln Tyr Tyr Leu Ala

20	25	30
----	----	----

Glu Pro Trp Gln Phe Ser Met Leu Ala Ala Tyr Met Phe Leu Leu Ile

35	40	45
----	----	----

Met Leu Gly Phe Pro Ile Asn Phe Leu Thr Leu Tyr Val Thr Val Gln

50	55	60
----	----	----

His Lys Lys Leu Arg Thr Pro Leu Asn Tyr Ile Leu Leu Asn Leu Ala

65	70	75	80
Val Ala Asp Leu Phe Met Val Phe Gly Gly Phe Thr Thr Thr Leu Tyr			
85	90	95	
Thr Ser Leu His Gly Tyr Phe Val Phe Gly Pro Thr Gly Cys Asn Leu			
100	105	110	
Glu Gly Phe Phe Ala Thr Leu Gly Gly Glu Ile Ala Leu Trp Ser Leu			
115	120	125	
Val Val Leu Ala Ile Glu Arg Tyr Val Val Val Cys Lys Pro Met Ser			
130	135	140	
Asn Phe Arg Phe Gly Glu Asn His Ala Ile Met Gly Val Ala Phe Thr			
145	150	155	160
Trp Val Met Ala Leu Ala Cys Ala Ala Pro Pro Leu Val Gly Trp Ser			
165	170	175	
Arg Tyr Ile Pro Glu Gly Met Gln Cys Ser Cys Gly Ile Asp Tyr Tyr			
180	185	190	
Thr Pro His Glu Glu Thr Asn Asn Glu Ser Phe Val Ile Tyr Met Phe			
195	200	205	
Val Val His Phe Ile Ile Pro Leu Ile Val Ile Phe Phe Cys Tyr Gly			
210	215	220	
Gln Leu Val Phe Thr Val Lys Glu Ala Ala Ala Gln Gln Gln Glu Ser			
225	230	235	240
Ala Thr Thr Gln Lys Ala Glu Lys Glu Val Thr Arg Met Val Ile Ile			
245	250	255	
Met Val Ile Ala Phe Leu Ile Cys Trp Leu Pro Tyr Ala Gly Val Ala			
260	265	270	
Phe Tyr Ile Phe Thr His Gln Gly Ser Asp Phe Gly Pro Ile Phe Met			
275	280	285	
Thr Ile Pro Ala Phe Phe Ala Lys Thr Ser Ala Val Tyr Asn Pro Val			
290	295	300	

Ile Tyr Ile Met Met Asn Lys Gln Phe Arg Asn Cys Met Val Thr Thr
305 310 315 320
Leu Cys Cys Gly Lys Asn Pro Leu Gly Asp Asp Glu Ala Ser Thr Thr
325 330 335
Val Ser Lys Thr Glu Thr Ser Gln Val Ala Pro Ala
340 345

<210> 3

<211> 223

<212> PRT

<213> Bos taurus

<220>

<223>

<400> 3

Ile Val Gly Gly Tyr Thr Cys Gly Ala Asn Thr Val Pro Tyr Gln Val
1 5 10 15
Ser Leu Asn Ser Gly Tyr His Phe Cys Gly Gly Ser Leu Ile Asn Ser
20 25 30
Gln Trp Val Val Ser Ala Ala His Cys Tyr Lys Ser Gly Ile Gln Val
35 40 45
Arg Leu Gly Glu Asp Asn Ile Asn Val Val Glu Gly Asn Glu Gln Phe
50 55 60
Ile Ser Ala Ser Lys Ser Ile Val His Pro Ser Tyr Asn Ser Asn Thr
65 70 75 80
Leu Asn Asn Asp Ile Met Leu Ile Lys Leu Lys Ser Ala Ala Ser Leu
85 90 95
Asn Ser Arg Val Ala Ser Ile Ser Leu Pro Thr Ser Cys Ala Ser Ala
100 105 110

Gly Thr Gln Cys Leu Ile Ser Gly Trp Gly Asn Thr Lys Ser Ser Gly
115 120 125
Thr Ser Tyr Pro Asp Val Leu Lys Cys Leu Lys Ala Pro Ile Leu Ser
130 135 140
Asp Ser Ser Cys Lys Ser Ala Tyr Pro Gly Gln Ile Thr Ser Asn Met
145 150 155 160
Phe Cys Ala Gly Tyr Leu Glu Gly Gly Lys Asp Ser Cys Gln Gly Asp
165 170 175
Ser Gly Gly Pro Val Val Cys Ser Gly Lys Leu Gln Gly Ile Val Ser
180 185 190
Trp Gly Ser Gly Cys Ala Gln Lys Asn Lys Pro Gly Val Tyr Thr Lys
195 200 205
Val Cys Asn Tyr Val Ser Trp Ile Lys Gln Thr Ile Ala Ser Asn
210 215 220

<210> 4

<211> 58

<212> PRT

<213> Bos taurus

<220>

<223>

<400> 4

Arg Pro Asp Phe Cys Lle Glu Pro Pro Tyr Thr Gly Pro Cys Lys Ala
1 5 10 15

Arg Ile Ile Arg Tyr Phe Tyr Asn Ala Lys Ala Gly Leu Cys Gln Thr
20 25 30

Phe Val Tyr Gly Gly Cys Arg Ala Lys Arg Asn Asn Phe Lys Ser Ala
35 40 45

Glu Asp Cys Met Arg Thr Cys Gly Gly Ala

50

55

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP02/00286

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl' G06F17/30, G06F17/50, G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl' G06F17/30, G06F17/50, G06F19/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1926-1996	Jitsuyo Shinan Toroku Koho	1996-2002
Kokai Jitsuyo Shinan Koho	1971-2002	Toroku Jitsuyo Shinan Koho	1994-2002

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Nobuhiro GO, "Amino San Chikan niyori Rittai Kozo Henka oyobi Sodo Tanpakushitsu no Rittai Kozo no Yosoku Gijutsu no Kaihatsu", Kinosei Tanpakusitsu no Kaiseki. Shushoku-Mogi Gijutsu no Kaihatsu ni Kansuru Kenkyu (Dai 2ki Showa 61 to 63nendo) Seika Hokokusho, 1989.08, pages 89 to 101, particularly, page 91, line 8 to page 92, line 1; page 93, lines 20 to 26	1-5
A	Nobuhiro GO, "II. Computer de Tanpakushitsu Bunshi o Miru", Japanese Scientific Monthly, 15 September, 1988 (15.09.88), Vol.41, No.9, pages 716 to 723; particularly, page 717, right column, line 26 to page 718, left column, line 30; Figs. 3 to 5	1-5

Further documents are listed in the continuation of Box C. See patent family annex.

- * Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search 05 April, 2002 (05.04.02)	Date of mailing of the international search report 16 April, 2002 (16.04.02)
Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP02/00286

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PAREY, M. et al., The Particle Concept: Placing Discrete Water Molecules During Protein-Ligand Docking Predictions. PROTEINS: Structure, Function, and Genetics 01 January, 1999 (01.01.99), Vol.34, No.1, pages 17 to 28, whole document	6-9
X	WO, 96/13785, A1 (Shoko ITAI), 09 May, 1996 (09.05.96), Claim 1 & EP 790567 A1 & US 20010018682 A1 & KR 97707500 A	12

INTERNATIONAL SEARCH REPORT

International application No.

PCT / JP02/00286

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: 10, 11
because they relate to subject matter not required to be searched by this Authority, namely:
(See extra sheet.)

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)This International Searching Authority found multiple inventions in this international application, as follows:
(See extra sheet.)

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP02/00286

Continuation of Box No.I of continuation of first sheet (1)

The "recording medium readable with a computer characterized in that atomic coordinates specifying a three-dimensional structure are described therein" as set forth in claim 10 and the "data base characterized by containing atomic coordinates specifying a three-dimensional structure" as set forth in claim 11 correspond to mere offer of information and thus relate to a subject matter which this International Searching Authority is not required, under Rule 42(4) of the Regulations under the PCT, to search.

Continuation of Box No.II of continuation of first sheet(1)

Claims 1 to 5 relate to the construction of a plural number of receptor protein models considering the heat wobble of molecules based on a normal mode of vibration.

Claims 6 to 9 relate to the specification of ligand-binding sites considering hydrophobic interactions in aqueous solutions.

Claim 12 relates to a method of designing a drug molecule with the use of data base.

These three groups of inventions are regarded neither as a single invention nor as a group of inventions so linked as to form a single general inventive concept. Such being the case, the present international application fails to satisfy the requirement of unity of invention.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int Cl' G 06 F 17/30, G 06 F 17/50, G 06 F 19/00

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int Cl' G 06 F 17/30, G 06 F 17/50, G 06 F 19/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1926-1996年
日本国公開実用新案公報	1971-2002年
日本国実用新案登録公報	1996-2002年
日本国登録実用新案公報	1994-2002年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JICSTファイル（JOIS）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	郷信宏「アミノ酸置換による立体構造変化および相同蛋白質の立体構造の予測技術の開発」機能性タンパク質の解析・修飾・模擬技術の開発に関する研究(第2期昭和61-63年度)成果報告書, 1989. 08, p. 89-101, 特に、p. 91第8行-p. 92第1行, p. 93第20-26行	1-5
A	郷信宏「II. コンピュータで蛋白質分子を見る」学術月報, 1988. 09. 15, Vol. 41, No. 9, p. 716-723, 特に、p. 717右欄第26行-p. 718左欄第30行, 図3-5	1-5

 C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

- 「A」特に関連のある文献ではなく、一般的技術水準を示すもの
- 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
- 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
- 「O」口頭による開示、使用、展示等に言及する文献
- 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」同一パテントファミリー文献

国際調査を完了した日 05.04.02	国際調査報告の発送日 16.04.02
国際調査機関の名称及びあて先 日本国特許庁（ISA/JP） 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員） 高瀬 勤 電話番号 03-3581-1101 内線 3599 5M 9069

国際調査報告

国際出願番号 PCT/JP02/00286

C(続き)	関連すると認められる文献	
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	PAREY, M. et al. The Particle Concept:Placing Discrete Water Molecules During Protein-Ligand Docking Predictions. PROTEINS:Structure, Function, and Genetics 1999.01.01, Vol. 34, No. 1, p. 17-28, Whole document	6-9
X	WO 96/13785 A1(板井昭子)1996.05.09 請求の範囲1 &EP 790567 A1 &US 20010018682 A1 &KR 97707500 A	12

第I欄 請求の範囲の一部の調査ができないときの意見(第1ページの2の続き)

法第8条第3項(PCT17条(2)(a))の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. 請求の範囲 10, 11 は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、

請求の範囲10の「立体構造を規定する原子座標が記載されていることを特徴とするコンピュータ読みとり可能な記録媒体」、請求の範囲11の「立体構造を規定する原子座標を含むことを特徴とするデータベース」は情報の単なる提示であり、特許協力条約に基づく国際出願等に関する法施行規則第42条(4)に定められた国際調査を要しない対象に係るものである。

2. 請求の範囲 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、

3. 請求の範囲 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第II欄 発明の單一性が欠如しているときの意見(第1ページの3の続き)

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

請求の範囲1-5は、基準振動モードに基づいて分子の熱揺らぎを考慮した複数の受容体タンパク質モデルの構築に関するものである。

請求の範囲6-9は、水溶液中での疎水相互作用を考慮したリガンド結合部位の特定に関するものである。

請求の範囲12は、データベースを用いた薬物分子設計方法に関するものである。

これら三の発明群が一の発明であるとも、單一の一般的発明概念を形成するように連関している一群の発明であるとも認められないから、この国際出願は発明の單一性の要件を満たしていない。

1. 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。

2. 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。

3. 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかつたので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。

4. 出願人が必要な追加調査手数料を期間内に納付しなかつたので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

追加調査手数料の納付と共に出願人から異議申立てがあった。

追加調査手数料の納付と共に出願人から異議申立てがなかつた。

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record.**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)